24166

National Library of Canada
Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* INGRAM John C.L.

TITLE OF THESIS/*TITRE DE LA THÈSE* Perceptual Dimensions of Phonemic Recognition

UNIVERSITY/*UNIVERSITÉ* University of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED/*GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* Ph. D.

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1975

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* Prof. W.J. Baker

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* 10/2/75 SIGNED/*SIGNÉ*

PERMANENT ADDRESS/*RÉSIDENCE FIXE*
28 KEEGANO
MILLWOODS
EDMONTON, ALBERTA

NL-91 (3-74)

THE UNIVERSITY OF ALBERTA

PERCEPTUAL DIMENSIONS OF PHONEMIC RECOGNITION

by

© JOHN C. L. INGRAM

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY
IN
PSYCHOLINGUISTICS

DEPARTMENT OF LINGUISTICS

EDMONTON, ALBERTA
SPRING, 1975

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "Perceptual Dimensions of Phonemic Recognition", submitted by John C.L. Ingram in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psycholinguistics.

Wm. J. Baker
Supervisor

Bruce L. Derwing

A. Rozsypal

Charles M. Bowman

Peter Ladefoged
External Examiner

Date 10th Feb. 1975

for Deloris

ABSTRACT

This study is an experimental inquiry into the perceptual dimensions involved in the recognition of a selected set of English consonantal phonemes. The methodology and substantive findings of previous attempts to determine the perceptual reality of phonological features, or to provide a direct characterization of perceptual features without recourse to a specific formal linguistic framework, are reviewed. Particular attention is given to multidimensional scaling (MDS) as a methodological tool and to the results of MDS of minimal speech sounds.

Four experiments were performed with related sets of English CV syllables, employing MDS and factor analysis of direct and indirectly elicited judgements of perceptual similarity. Analysis of the data yielded two major dimensions, each with a ready auditory/acoustic characterization. Various phonological feature systems and measurable acoustic properties of the experimental stimuli were used as predictors of both the raw perceptual proximity matrix and the set of MDS derived inter-stimulus distances. It was found that the MDS configuration could be reconstructed with a fairly high degree of accurracy (Pearson's Product Moment Correlation = .81) from the physical duration of the consonantal portion of the stimulus and a simple bandfilter function, labelled the Resonance-Hiss dimension. Theoretical implications of the findings for a model of perception at the phonemic level are discussed.

ACKNOWLEDGMENTS

I would particularly like to thank Prof. William Baker my thesis supervisor for the help and advice that he has provided at all stages in the writing of this thesis. His contribution ranges from matters of general conceptual framework, to sharpening many vaguaries of thought and style that might otherwise have found their way into the final version.

I am grateful also for the contributions made by other members of my committee. Dr. Bruce Derwing provided valuable suggestions, particularly on questions of phonological theory. I am indebted to Dr. Anton Rozsypal for substantial technical assistance, for bringing some important work to my attention, and for the lively interest he has shown in this project. I would like to extend thanks to Dr. Bourassa and Prof. Peter Ladefoged who kindly consented to serve on my examining committee. Ladefoged's work is a source of inspiration for many students of phonetics but meeting him in person adds uniquely to one's enthusiasm for the subject.

To the many students who volunteered their services as subjects for the experimental work I would like to give special thanks, and thanks also to their instructors who generously contributed class time. In particular I wish to thank Mike Esgrow, Bob Atkinson, Peter Boothroyd, Anne

Lambert and Stephen Scobie. Whatever doubts they may have had, or still do, about the value of this kind of pursuit, they had the charity to set aside, for which I am appropriately grateful.

Finally, I would like to thank Prof. Ian Stuart for the interest he has shown in my work, and for a general conception of the goals of linguistic research that initially attracted me to the discipline and which I continue to find exciting.

This work was supported by a teaching assistantship from the Department of Linguistics and a dissertation fellowship from The University of Alberta.

TABLE OF CONTENTS

CHAPTER PAGE

I INTRODUCTION 1

The Distinctive Features Model 4

Controversial Issues in Speech Recognition 10

Methodology 14

II PHONOLOGICAL THEORY IN RELATION TO A PERCEPTUAL MODEL 20

Features 21

Jakobsonian Features 22

The Binary Principle 25

Jakobson Summarised 30

Features in Generative Phonology 31

III PERCEPTUAL DIMENSIONS IN PHONEMIC RECOGNITION 40

Confusion Matrices 44

Questions posed by the Miller and Nicely study ... 50

Perceptual Confusions under non-noisy conditions . 54

Perceptual Proximities via Similarity Scaling 59

Conclusions 69

IV ELEMENTS OF MULTIDIMENSIONAL SCALING 73

Estimation of Dimensionality 77

Measurement Error and Configuration Recoverability 78

The Problem of Rotation and the INDSCAL model 80

The Spatial Metric 84

V EXPERIMENTAL RESULTS 88

Experiment I 90

Choice of Stimuli 91

Method 91

Procedure 93

Scoring the Responses 94

Results 95

Experiment II 99

Results 101

Experiment III 107

Results 108

Experiment IV 109

Method 111

Results 112

VI ANALYSIS AND DISCUSSION 117

Regression Analyses of Interpont Distances and Proximity Scores 120

Perceptual Rating Scales as Prediction Variables 127

Physical Correlates of the Hypothesised Perceptual Dimensions 129

Broader Discussion of Findings 142

VII CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH 149

REFERENCES 156

APPENDIX A: Distinctive Feature Definitions 163

APPENDIX B: Kruskal Scaling Experiment I 165

APPENDIX C: Proximity Matrices, Stress Plots, Experiment II 166

APPENDIX D: Kruskal Scaling Experiment II 167

APPENDIX E: Torgerson Scaling Experiment II 168

APPENDIX F: Factor Analysis Experiment II 172

APPENDIX G: Factor Analysis Experiment IV 176

APPENDIX H: Phonological Features: Regression Analyses .177

APPENDIX I: Rating Scales: Regression Analyses 179

APPENDIX J: Oscillograms of test stimuli 180

APPENDIX K: Acoustic Properties: Regression Analyses ...181

APPENDIX L: Bandfilter Functions for 12 Experimental Stimuli 182

LIST OF TABLES

Table Description Page

3.1 Details of the Miller and Nicely Experiment 44

3.2 Details of the Wang and Bilger Experiment 51

3.3 Perceptual Salience of Features - White-Noise Condition (Wang & Bilger, 1973) 53

3.4 Perceptual Salience of Features - "Quiet" Condition (Wang & Bilger, 1973) 56

3.5 Prominence of Duration as a function of S/N Ratio (Wang & Bilger, 1973) 58

3.6 Prominence of Voicing as a function of S/N Ratio (Wang & Bilger, 1973) 59

3.7 INDSCAL Dimensions and Weightings for Three Scaling Methods (Singh, Woods, & Becker, 1972) ... 68

5.1 Proximity Matrix Experiment I 95

5.2 Average Rank Scores on Rating Scales Experiment IV 113

5.3 Rank Ordered Estimated Scale Reliabilities Experiment IV 114

6.1 Distinctive Feature Systems 123

6.2 Stimulus Loadings on Rotated Kruskal Dimensions 131

6.3 Physical Predictors of "Resonance-Hiss" Dimension Loadings 135

6.4 Correlations Between Predictor Variables 137

LIST OF FIGURES

Figure Description Page

3.1 Reanalysis of Miller and Nicely (1955) 48

3.2 Reanalysis of Graham and House (1971) 57

3.3 Hierachical Clustering Analysis: Shepard (1972) 61

3.4 Reanalysis of Black (1968) 65

3.5 Reanalysis of Singh, Woods, and Becker (1972) 69

5.1 Stress X Dimensionality Experiment I 96

5.2 Three-Dimensional Kruskal Scaling Experiment I 97

5.3 Two-Dimensional Kruskal Scaling Experiment I 100

5.4 Two-Dimensional Kruskal Scaling Experiment II 102

5.5 Torgerson Scaling and Factor Analysis Experiment II 106

5.6 Two-Dimensional Kruskal Scaling Experiment III 110

5.7 Factor Analysis of Phone Rating Data Experiment IV 115

6.1 Stepwise Regression Analysis: Distinctive Features as Predictors of Derived Distances and Raw Proximities 125

6.2 Stepwise Regression Analysis: 13 Perceptual Scales as Predictors of Derived Distances and Raw Proximities 128

6.3 Instrumentation for Acoustic Analysis of "Resonant-Hiss" Dimension 134

6.4 Reconstruction of Two-Dimensional Perceptual Configuration on basis of Physical Variables 138

6.5 First Eigenvector of the Variance-Covariance Matrix of Bandfilter Spectra (Pols, 1974) 139

CHAPTER I

INTRODUCTION

Although it is just one facet of the problem of speech perception, phonemic recognition denotes a perceptual capability that is arguably fundamental for any general model of how the listener extracts linguistic messages from the highly variable signal of human speech. It is fundamental in the sense that higher order syntacto-semantic elements of the message are predicated upon the extraction of a certain (unspecified but necessary) amount of phonological information from the signal. Some qualification may be in order here.

In all probability speech perception is a multilevel process. Superimposed upon the basic information flow from concrete auditory perceptual targets to abstract conceptual elements of the message, there appears to be a substantial amount of independent parallel processing of the signal at different levels of analysis. Decoding at the phonological level is no doubt partially directed by the listener's expectations or anticipations about the content of the message formed on the basis of ongoing semantic and syntactic processing. Phonological decoding is also most likely facilitated by the listener's knowledge of segmental and sequential redundancies that form the characteristic sound pattern of his native language. (The efficacy of this latter class of factors is attested by the listener's characteristic insensitivity to phonetic variation from the phonotactic constraints of his language.)

Despite the obvious fact that no particular level of signal analysis is functionally independent of any other in the perceptual process, (and indeed the isolation of any particular level of analysis seems to involve a degree of imposition of an artificial conceptual framework on the phenomenon under study), a case can be made for considering phonemic recognition as a significant and isolatable level in speech perception. Native listeners are remarkably consistent in their ability to extract sequences of phonemic targets from the quasi-continuous and highly variable speech signal. As Smith (1973) and others have observed, this achievement is of comparable complexity to the attainment of object constancy in vision, despite the instability of the pattern of retinal stimulation.

Several sources of complexity in the mapping between the acoustic signal and the stable phonemic target are identifiable: structural differences between the vocal tracts of different speakers; idiolectical variation arising from idiosyncracies in a speaker's manner of speech production; dialectical variation; coarticulation effects and other variations in the phonetic realization of a phonemic target associated with the linguistic environment in which the target is embedded; the background listening conditions themselves.

Although something of a linguistic trusim, the notion that the inventory of phonemic targets embodies the set of minimal meaning differentiating sound contrasts for a particular language, is also an important consideration for any model of speech perception. An efficient lexical storage code (and some kind of lexical storage system is an obviously necessary component of any model of perception or production) must in some way utilize this set of contrasts. It is, of course, conceivable that each lexical item could be recognized or reproduced on the basis of feature specifications unique to that item (the maximally inefficient option), but this would raise difficulties explaining how listeners can readily reproduce (through mimicry or orthographic transcription) phonemic sequences that, in all likelihood, they have never heard before. Also, as has often been remarked (e.g., Schane, 1973) it would be difficult to account for speakers' intuitions about the identity or contrast of particular phonemic targets in different linguistic environments if the psychological reality of the phonemic level of representation is denied.

However, it has often been argued to the contrary that phonemic recognition is more correctly regarded as a cognitive rather than a perceptual skill and that it is derivatively based on some more strictly perceptual unit of signal analysis. Massaro (1972) presents a good deal of evidence (not all of it sound) "in support of the interpretation... that the phoneme is not perceived directly...but is inferred from the identification of the syllable or word." Massaro takes the classical spectrographic evidence from synthetic stop consonants that Liberman et al. (1957,1967) used to support the parallel decoding of target phonemes upon segments of the acoustic signal of the order of the syllable as "... convincing support for eliminating the phoneme as the perceptual unit for processing speech." At the present time it is necessary to leave the question of the appropriate units of analysis open. Suffice to point out that what can be regarded as phonemic recognition can also be satisfactorally described in either syllabic, segmental, or subphonemic terms.

The Distinctive Features Model

Knowledge of the perceptual processes underlying phonemic recognition is at present very slight. Perhaps the most influential source of theoretical constructs in the past 20 years has been the Jakobson, Fant, and Halle monograph *Preliminaries to Speech Analysis* (1951).

Essentially, their model postulates a highly restricted, language universal, set of binary perceptual "features", each with a more or less clearly defined acoustic referent. Target phonemes are presumably mutually discriminated and identified by the speech processor on the basis of an array or pattern of "on-off" feature detector output states. The feature detectors are generally conceived to operate independently and in parallel upon the input signal (or more precisly, upon an appropriately speaker and speech rate normalized transform of the original signal). The authors draw attention to the fact that in the speech signal there is really no simple linear ordering of acoustic cues corresponding to a feature specification of the sequence of phonemes given by a phonological representation of the message.

A phonemic target is formally defined by an array of feature values. But the feature specification is not invariant over different realizations of the target in the speech signal. Phonological processes (such as assimilation and neutralization) are largely responsible for this variability. For correct phonemic recognition in many instances, the feature detector would have to take account of features of immediatly surrounding phonemic targets and "know" the relevant sequential constraints on feature combination for the language in question. This complicates but does not invalidate the simple parallel feature extraction model previously outlined.

More serious consequences flow from the observation (implied if not explicitly stated in the Jakobson, Fant, and Halle account) that the _perceptually_ relevant signal properties for recognizing a given feature are not necessarily those acoustic properties associated with the definition of that feature. Aspiration, for example, is in all likelihood the most salient auditory cue for the Tense-Lax feature in English stops, word or syllable initially. However, in word or syllable final position the "redundant" feature of the length of the immediatly preceding vowel seems to be the salient cue (Denes, 1955; Peterson & Lehiste, 1960; Raphael, 1971). Apart from the unfortunate semantic consequence of requiring "redundant" or "predictable" phonetic features to serve as the basis for inferring the presence of otherwise imperceptable "distinctive" features, the motivation for postulating distinctive features in a perceptual model is considerably weakened. As Smith (1973) has commented:

> Superficially at least, the theoretical advantages of this (the distinctive feature) approach are enormous. Rather than store the vast range of possible realizations ...of the phoneme /p/ and discriminate such patterns from a large number of alternative patterns, the perceiver needs only to keep track of a small number of features, each feature limited in the number of feature values it can adopt [p.512].

But at least some of the features in the Jakobson, Fant, and Halle system must be of comparable complexity with the phonemic targets themselves, in terms of their mapping onto perceptually salient characteristics of the auditory signal. If this is so, then the question that naturally arises is, what does a model of perception gain by postulating such entities? Also, doubt is cast on the perceptual reality of the features themselves. (In what sense is the "feature" which distinguishes /i/ and /I/ the same as that which differentiates /p/ from /b/ and /s/ from /z/?) For independent reasons, one may wish to retain features such as "gravity" and "tenseness" in a performance model. However, it should be recognized that doing so will probably complicate rather than simplify the account of the perceptual process.

The question of the relationship between phonological theory and a theory of speech perception at the phonological level will be dealt with more fully below (Chapter II). Suffice for present purposes to note the change in connotation that the term "distinctive feature" is apt to undergo as the topic shifts from phonological theory to speech recognition. In the former context distinctive features are primarily classificatory devices, conceived by phonologists, for the purpose of grouping segments that undergo similar phonological processes in identical linguistic environments. According to the celebrated "simplicity metric" of Halle (1964), that feature or feature system which most economically designates such "natural" classes of segments is most highly valued. In the context of a perceptual model however, a "distinctive feature" refers to some attribute (simple or complex from the viewpoint of a physical description of the signal) that serves for the native listener to differentiate between members of a set of perceptual targets. It is a common but dangerous assumption to treat these two senses of the term "distinctive feature" as synonymous. Thus Schane (1973) suggests: "The more features operating to distinguish segments, the more different those segments are and the greater their perceptual distance." The features to which Schane is referring are those of the M.I.T. School of generative phonology. While there does appear to be some positive relationship between most distinctive feature counts and the perceptual distance between two target phonemes (Smith, 1973; Singh, 1974), there are no good grounds for assuming that an optimal set of classificatory features for a generative phonology will optimally predict perceptual contrasts.

Interest in this paper focuses upon developing a feature system that is optimal from the perceptual point of view. Such a system must be capable of representing the entire set of phonemic contrasts that the hearer is accustomed to making and which are characteristic of his native language. More than this, however, a perceptually based feature system should reflect the relative perceptual salience of different phonemic contrasts for the phonetically untrained listener. It is intuitively evident that certain phonemically relevant sound contrasts are more easily detected by the auditory system than others. Such contrasts are probably acquired earlier in the course of language development and, it is reasonable to speculate, will be utilized in a greater proportion of the world's languages. The isolation of maximally perceptually contrastive phonemic classes - particularly if such classes should prove universally marked in human languages and ontogenically prior to other, more language specific, phonemic contrasts - would obviously be very important for the development of a phonological theory that seeks to attain a level of "explanatory" adequacy. Moreover, it is difficult to see how, within the confines of current methodology in phonology, such perceptual classes or sound groupings could be identified, if for no other reason than production and perception considerations are thoroughly confounded.

A perceptually based feature system should also specify what auditory properties of the signal are used in phonemic recognition, and thereby indicate something of the nature of the sound-analysing devices employed by the human perceptual apparatus. It is commonly agreed (and also implied in all of the proposed distinctive feature systems) that the nature of the auditory stimulus for phonemic recognition is multidimensional. But there is no substantial agreement about the number or the nature of these supposed perceptual dimensions. There is, correspondingly, little agreement about the general properties of the sound-analysing devices required for phonemic recognition.

Controversial Issues in Speech Recognition

Those associated with one highly influential trend in speech perception research, centred around the pioneering work of Liberman, Cooper, and their associates at Haskins Laboratories, have argued for the unique character of speech recognition as distinct from other forms of auditory recognition. They have postulated signal analysing devices (conceived of as specific linguistic feature detectors) that uniquely function for the recognition of speech. Studies under the dichotic listening paradigm (Kimura,1961; Studdert-Kennedy & Shankweiler, 1970; Darwin, 1971; Day & Bartlett, 1971; - to mention only a notable few), and a couple of recent experiments with Auditory Evoked Responses (Wood, Goff, & Day, 1971; Wood, 1973), it is claimed, support a neurophysiological locus for these perceptual devices in the left temporal hemisphere. It has also been suggested, on the basis of Voice Onset Time studies (Eimas et al., 1971) that these devices are "prewired" as part of the genetically transmitted neurological substrate of human linguistic capabilities. Such a notion is compatable with Jakobson's idea that there exists a relatively small set of features for constructing phonemic oppositions, a subset of which is realized in any given language.

We may contrast this view with the one (perhaps best typified by Lane (1965)) which argues that the stimulus dimensions employed, and the processes underlying phonemic recognition, are not uniquely linguistic (part of some innate human "faculte de langage") but ultimately derive from general properties of the auditory-perceptual system in interaction with an equally general process of discrimination learning. There may be selective "sharpening" or "suppression" of many potentially audible characteristics of the signal when it is processed as speech, but (the argument runs) there is no need to invoke special linguistic features and their corresponding detection devices, to explain phonemic recognition.

These two conflicting theoretical dispositions are useful for characterising "the state of the art" in speech perception research and for pointing up major specific problem areas. Individual writers can be roughly located on this theoretical continuum, ranging from strong "nativist" to strong "empiricist" positions.

The nativist position is characterised by strong assumptions about the nature of the user's language code and strong emphasis upon the differences between speech recognition and other modes of auditory perception. Proponents of this view have a tendency to accept the rubric of modern phonological theory (in particular, the generative phonology of Jakobson, Halle, Chomsky, Postal, et al.) for what it purports to be - a competence model: or an abstract representation of what the speaker-hearer "knows" about the sound pattern of his language.

The "empiricist" position makes minimal assumptions about the nature of the code and emphasises the continuity between speech perception (at the level of phonemic recognition) and what is known about the process of auditory perception in general. Proponents of this view will tend to reject linguistically derived constructs such as "phonemes" or "distinctive features" as components of a perceptual model - much less accept that the brain has evolved special perceptual decoding mechanisms to detect such entities in the speech signal. The major dimensions involved in speech perception should be predictable from the acoustic properties of the signal, the response characteristics of the auditory system, and the language experience of the listener.

Present knowledge is quite inadequate for choosing between these two broadly sketched alternatives. With respect to the "empiricist" view, it may be observed that there is insufficient understanding of the human auditory system to specify what major auditory parameters would be used for differentiating signals with roughly the same source characteristics as human speech. The great developments in instrumentation for acoustic research over the last 25 years have, in this respect, merely served to focus more sharply the problem of choosing a perceptually revealing physical representation of the signal.

With respect to the "nativist" view and the problem of characterising the language user's code - it is clear that this is very much an open question. The grammarian's formal constraints on rule writing (insofar as such agreed upon principles exist in this controversial area) have no clear relevance or plausability for a model of what the user "knows" (consciously or tacitly) about his language, as revealed in language use. For example, many postulated phonological processes reveal morphological relationships which permit a good deal of simplification and size reduction of the lexicon. But the psychological implication, that the resulting savings of lexical storage "space" are relevant for the language user, is quite unfounded (see Derwing, 1973, p.154, note 2).

Obviously not all the theoretically interesting questions which are raised by the two conflicing dispositions outlined above will be answerable in the forseeable future. However, what - with an admittedly uneasy choice of terminology - may be labelled the "nativist-empiricist" controversy in speech perception, points to two potentially fruitful hypotheses about the nature of the perceptual dimensions involved in phonemic recognition:

H1: The perceptual dimensions are unique to sounds of speech and presuppose a special "speech mode" of perceptual processing. Such dimensions may be either language specific (reflecting the particular system of phonological contrasts of a particular language) or universal (reflecting the set of "possible" phonemic oppositions).

H2: The perceptual dimensions used in phonemic recognition are not speech or language specific but general auditory properties ultimately attributable to the way the auditory apparatus responds to complex environmental sounds with acoustic properties roughly similar to the human speech signal. (A necessary qualification on this hypothesis should probably be made to the effect that any "general auditory parameters" will be modified or "tuned" by specific learning experience and the demands imposed by the sound contrasts of a given language.)

The methodological question which is prior to these hypotheses, of how such perceptual dimensions are to be isolated and described, is taken up in the following section.

Methodology

At the risk of oversimplification, two major research strategies to study the problem of phonemic recognition are discernable. These two strategies are respectively associated with two types of difficulty confronting the researcher - the twin horns of the speech perception dilemma

- indeterminancy with respect to the description of the signal and indeterminancy with respect to the description of the code.

The first research strategy, which forms the dominant paradigm in contemporary phonetics, directly addresses the first horn of the dilemma. By controlled manipulation of physical parameters of the signal the investigator attempts to isolate those features of the signal that are relevant for some specific phonemic distinction or class of distinctions. The power of this paradigm has only been realised with the development of speech synthesis (begining with the early Pattern Playback method) and computer based techniques of signal manipulation and reconstruction (Roszypal, 1974). However, this approach has real limitations from the viewpoint of the second horn of the dilemma. The perceptual importance of the linguistic distinction under investigation must be taken as given. The now extensive body of research on the topic of voice onset time (VOT) is a convenient example here. VOT is just one of a number, and probably not the most salient, of acoustic cues for the recognition of voicing, and the perceptual status of the voice feature is itself problematical. What, for example, is its relative prominance compared with other phonemic distinctions such as stridency? What justification is there, from the perceptual point of view, for regarding voicing as a homogeneous feature, applicable right across the phonemic inventory?

In short, the dominant paradigm of experimental phonetics yields a great deal of precise information about properties of the signal associated with specific phonological distinctions deemed relevant to phonemic recognition. But it is difficult to decide from all this information, what properties of the signal are of greater or lesser importance for the perceptual system, and how the collective findings of a large number of contextually restricted experiments can be integrated into a general model of phonemic recognition.

The second paradigm (which can be ascribed to the psychologists, having generously yielded the former to the phoneticians) is focused less upon the signal than the linguistic code itself, or more precisely, upon the problem of determining an adequate representation of speech sounds as perceptual end-products of the decoding process. It begins with a consideration of the total set of phonemic distinctions the listener is normally capable of making. This is embodied in the full phonemic inventory of the listener's native language and the researcher who opts for this paradigm attempts to map the perceptual relationships that obtain amongst the set of target phonemes. Such a mapping will yield the perceptually most salient contrasts amongst the set of target phonemes, from which it should be possible to draw inferences about what features of the signal are most important for phonemic recognition and perhaps too, broad suggestions as to how the decoding takes place. Hence, although the two paradigms outlined above are mutually complementary, it seems that the second has a certain logical priority.

The first step in mapping the perceptual relationships among a set of target phonemes is to obtain a matrix of proximities - a convenient numerical representation of the degree of perceptual relatedness between any given target and all other targets in the set. A proximity matrix is a set of experimentally generated estimates which serves as the data base for, or the empirical constraint upon, the investigator's attempt to model the output of the perceptual process. Phonemic targets that share a common perceptual basis will have high valued entries in the proximity matrix. The most perceptually contrastive will have the lowest entries. The matrix is symetrical about the main diagonal where the values (which represent the degree of relatedness between a target and itself) are maximal.

A variety of experimental methods have been developed for generating perceptual proximity matrices. These differences may arise from the scaling procedures or, more importantly, from the kind of data base that the experimenter chooses to adopt. This methodological variation greatly complicates any review of the reported research, because different methods of generating a proximity matrix may have different implications for the study of speech perception. For example, Wickelgren's (1966) oft quoted study of phonemic substitutions is of much greater relevance for processes involving the retention of phonological information (say, for purposes of lexical retrieval) than for first or second order phonological decoding. Chapter III of this paper discusses the various methods of generating proximity matrices that have been proposed, compares their differential theoretical implications and reviews the major findings reported to date.

Obtaining a proximity matrix is only the initial step in mapping perceptual relationships among a set of perceptual targets. A variety of powerful data reduction techniques have been developed for characterising the presumed latent structure in a proximity matrix. Each of these data reduction techniques imposes a given mathematical model upon the data. The question of the appropriatness of particular formal structures for a perceptual model therefore becomes a critical question, but one very difficult to answer.

Usually the latent structure of a proximity matrix is characterised upon the assumption that proximities may be interpreted as distances (or, more precisely, as some function of the unknown distances) in a space of certain specified metric properties and dimensionality. A family of data reduction procedures known as "multidimensional scaling" (MDS) has been developed over the last 20 years (primarily in the last 10) for obtaining such spatial representations of the proximity matrix. Also relevant are the related methods of "factor analysis." Chapter IV deals with the basic procedures and foundational assumptions of MDS as it applies to the study of speech perception.

Succeeding chapters describe a series of experiments with selected sets of English consonantal phonemes embedded in a CV frame, employing multidimensional scaling and factor analysis to subjects' direct and indirect judgements of perceptual similarity. The aim of all of these experiments was to attempt to determine the number and nature of the major perceptual dimensions native listeners employ in recognizing the consonantal phonemes of English.

Ancillary to this major theme are questions concerning:

(1) the relative perceptual prominance of various phonologically derived "distinctive features" and the perceptual adequacy of different distinctive feature systems.

(2) the perceptual and acoustic correlates of interpretive factors derived from the MDS studies.

Multiple Regression Analysis was used as the major analytical tool here. Both the derived interpoint distances and the raw proximity scores were used as dependent variables to which linear, least squares, fittings were obtained for selected sets of phonologically, perceptually, and acoustically based independent variables.

CHAPTER II

PHONOLOGICAL THEORY IN RELATION TO A PERCEPTUAL MODEL

In the previous chapter certain difficulties were noted when the Jakobson, Fant, and Halle distinctive feature theory is treated as a schematic outline for a model of phonemic recognition. This raised the wider question of the relevance of phonological theory for understanding how listeners extract linguistic information from the speech signal. This question is also urged by the claims generative phonologists make for their grammars as descriptions of "linguistic competence."

It is generally conceded that "the fundamental unit of generative phonology is the distinctive feature... [Harms, 1968]." Therefore, the strategy adopted in this chapter for clarifying this question of the relevance of phonological theory for the study of speech perception, will be to note what functions distinctive features serve and to inquire what constraints operate upon their postulation. On the basis of this analysis it should be possible to see how well the linguistic functions of distinctive features may subserve necessary or optionally useful functions of a speech recognition device.

Features

Distinctive features fulfill three basic functions in phonological theory. Firstly, they are used to specify phonetic properties, so that the set of feature values assigned to a segment can provide an overall index of phonetic similarity with any other segment and so that sounds can be compared with respect to particular phonetic qualities. Secondly, features have traditionally been employed to specify phonemic oppositions - those sound contrasts of a language that are of special linguistic significance over other perceivable auditory contrasts in the signal. Thirdly, distinctive features are used for grouping together segments that undergo the same phonological processes. These may be respectivly referred to as the phonetic, the phonemic, and the phonological (or classificatory) functions of distinctive features.

These functions are to some extent open to various interpretations, and different theorists do not necessarily ascribe to them the same relative importance. Phonetic relationships can be described with either production or perception primarily in mind. Perhaps the most obvious contrast between the Jakobsonian and the Chomsky and Halle feature systems involves just this orientation. For Jakobson, distinctive features were undoubtedly thought of as perceptual entities:

> It is the dichotomous scale of the

> distinctive features... that to a large extent determines our perception of the speech sounds [Jakobson, Fant, & Halle, 1951, p. 10].

> ... in its sound shape any language operates with discrete and polar distinctive features, and this polarity enables us to detect any feature functioning ceteris paribus [1957, p. 11].

For the generative phonologists, on the other hand, phonetic features are defined mainly in articulatory terms with no explicit claims regarding their perceptual reality:

> The total set of features is identical with the set of properties that can in principle be controled in speech; they represent the phonetic capabilities of man, and we would assume are therefore the same for all languages [Chomsky & Halle, 1968, p. 259].

> ... the phonetic matrix is then descriptive of the fact that the human vocal system is composed of a number of subparts capable of independent action and of different types of action... [Postal, 1968,p.59].

Any claim for the perceptual reality of these kinds of features is evidently extrinsic to the features themselves and must derive from a special hypothesis about the nature of speech perception, perhaps along the lines of the "motor theory" of perception of Liberman et al. (1967) or the "analysis by synthesis" model of Halle and Stevens (1967).

Jakobsonian Features

Although he thought of features primarily in auditory-perceptual terms, Jakobson was not concerned with a detailed auditory phonetic description of speech sounds. Distinctive features provided a vehicle for representing the set of phonemic oppositions of a language and Jakobson believed that it is these and not the "redundant" auditory-phonetic features of speech sounds to which the native speaker responds when he listens to speech.

In itself, the phonemic function imposes no significant constraints upon the choice of features. Jakobson constrained feature choice by requiring that they be universally applicable to all languages and small in number. He found he was able to considerably reduce the number of necessary features by applying the same set of features to capture both vowel and consonantal phonemic oppositions.

> For example, it can be shown that the relation of the close to open vowels, on the one hand, and that of the labials and dental consonants to consonants produced against the hard and soft palate, on the other, are all implementations of a single opposition: diffuse vs. compact... In their turn the relations between the back and front vowels, and between the labial and dental consonants pertain to a common opposition: grave vs. acute [Jakobson, Fant, & Halle, 1951, p. 7].

The rather bold innovation of cross-classifying both consonants and vowels with the same set of features has been largely retained by generative phonology - though not for the purpose of minimizing the number of features, but to provide a notation to capture the assimilatory nature of certain consonant-vowel interactions (such as palatalization or rounding). The phonological function of distinctive features was not explicitly formulated by Jakobson, but has been the major criterion in the later revision of his scheme by generative phonologists. Whether the features that cross-classify (as distinct from those that merely separate) consonants and vowels, make sense as perceptual dimensions, is seriously doubtful. Jakobson, Fant, and Halle (1951) discuss the acoustic and perceptual characteristics of the two major cross-classificatory features separately for consonants and vowels. This alone seems to indicate that there are problems with gravity and compactness as singular and coherent perceptual dimensions that apply right across the consonant-vowel domain.

Jakobson also economized on features by exploiting the fact that certain kinds of phonemic opposition never appear to co-occur in any one language. This enables superficially distinct oppositions to be collapsed into a single binary feature without restricting the range of phonemic oppositions that can be represented by the system. For example, the flat vs. plain feature is used for both phonemic labialization and pharyngealization (among others). Although from an articulatory standpoint it is difficult to imagine a more contrastive pair of articulatory gestures, Jakobson claims that their perceptual effects are so similar that no language could tolerate their coexistence as independent phonemic oppositions:

> The fact that peoples who have no pharyngealized consonants in their mother tongue, as, for instance, the Bantus and the Uzbeks, substitute labialized articulations for the corresponding pharyngealized consonants of Arabic words, illustrates the perceptual similarity of pharyngealization and lip rounding [1951, p. 31].

This example rather nicely points to the difference in orientation between the Jakobsonian and the Chomsky and Halle feature systems which was noted above. If features were in any sense to represent ideal targets or instructions for the articulatory apparatus, the collapsing of pharyngealization and labialization into a single flat vs. plain feature dimension would be clearly inappropriate. (Brain to Mouth: "Either purse your lips or constrict your pharynx". - cf., Mc Cawley, 1972).

The Binary Principle

Although it does not contribute to minimizing the number of features, but rather works to the contrary, Jakobson regarded the binary nature of distinctive features as fundamental:

> In the special case of speech... a set of binary selections is inherent in the communication process itself as a constraint imposed by the code on the participants in the speech event... The dichotomous scale is the pivotal principle of the linguistic structure. The code imposes it upon the sound [1951, p. 9].

On the basis of this and quotations made earlier one might be inclined to regard the binary principle as axiomatic and not as an empirical hypothesis. However, attempts are made to support the binary nature of distinctive features with psychological evidence (Jakobson and Halle, 1956), so this ambiguity, at least, is resolved.

Jakobson and Halle's first argument is that the auditory system is maximally efficient when operating with binary feature dimensions:

> Recent experiments have confirmed that multidimensional auditory displays are most easily learned and perceived when 'binary coded' [Jakobson & Halle, 1956, p. 47].

In fact, there is a dearth of relevant information on this admittedly important subject. The experiments to which Jakobson and Halle refer (Pollack & Ficks, 1954) are described by the authors themselves as "at best, only exploratory," and do not provide the right kind of information that is needed to test the binary coding hypothesis. Pollack and Ficks created multidimensional auditory stimuli of 6 and 8 dimensions using tone peeps and noise bursts. (The dimensions were created out of of such variables as the frequency of the tone, the rate of alternation of tone and noise, and the overall duration of the stimulus.) Two, three, or five steps were created for each stimulus dimension, with the two steps in the binary condition defined by the extreme steps on the five step scale and the three steps on the ternary condition by steps 1, 3, and 5 on the five step scale. All dimensions were binary for the eight-dimensional stimuli. Subjects worked with either binary, trinary, or quinary multidimensional stimuli. Their task was to identify the state of each stimulus dimension separately on any given stimulus presentation. Judgements were not made under time pressure. Neglecting details of the quantitative analysis, the authors' major results are of interest:

> The most striking finding is that the amount of information transmitted with these multidimensional auditory displays greatly exceeds that obtained under comparable conditions with unidimensional displays...
>
> The second finding is that there is proportionately little improvement in information transmission as each dimension is subdivided more finely... the more proficient subjects were able to take better advantage of the finer subdivision of each dimension...
>
> The third finding is that a further increase in the number of stimulus dimensions produced a still further gain in information transmission [1954, p. 156].

The second and third findings are hardly surprising considering: (a) that the most contrastive steps were selected as the values for the binary condition, (b) the nature of the subject's task, which was to make a series of successive judgements about the stimulus without time pressure. The information transmission measure thus overlooked "the most important single variable in information transmission: time." Consideration (a) shows that the design of the experiment clearly favours the binary displays. Consideration (b) shows that it favours increasing the simulus dimensionality. Arguably, what the experimenters should have done was to study the effect of stimulus dimensionality and dimensional subdivision upon the amount of information that can be transmitted per unit time. The authors of the experiment were aware of the limited generalizability of their results, but apparently Jakobson and Halle were not.

Jakobson and Halle's second piece of "evidence" I will have to simply quote because it is quite obscure with respect to both intrinsic meaning and empirical implication:

> Second, the phonemic code is acquired in the earliest years of childhood and, as psychology reveals, in a child's mind the pair is anterior to isolated objects (Wallon H., 1945). The binary opposition is a child's first logical operation. Both opposites arise simultaneously and force the infant to choose one and supress the other of two alternatives [Jakobson & Halle, 1956,p.47].

Their third and final piece of evidence, involving an early experiment with vowel mixing (Huber,1934), is too restrictive in scope to be worth considering in detail. Besides, as Jakobson and Halle admit, the results are equivocal. They disconfirm the binary hypothesis on the compact-diffuse dimension with English front vowels and (in

Jakobson and Halle's eyes, but not this author's), confirm it with the vowels /u/, and /i/ "on the tonality axis."

Perhaps the source of the binary principle lies in the traditional methodology of phonemic analysis which involves the examination of sets of minimal pair oppositions in an attempt to discern a number of elementary phonetic features that, by their presence or absence, can serve to subclassify phones into phonemic classes. The minimal pairs test may be thought to imply a binary opposition. However, this is of no theoretical significance, being merely a consequence of the analytical technique.

It is not disputed here that, in terms of subjective perceptual impression, speech sounds have a quantal character. Moreover, there is a considerable body of evidence based upon natural and synthetic speech (Liberman et al., 1961; 1963; 1970; Lane, 1965; 1967; for review articles) that for certain kinds of phonemic discrimination (most notably, involving the stop consonants) perception is categorical in the sense that the listener's discrimination sensitivity is not constant along the relevant phonetic dimension, but peaks at phoneme boundary points. However, to admit the quantal nature of the subjective percept and the categorical response of the human auditory system in some areas of phonemic recognition, in no way entails acceptance of the binary principle espoused most strongly by Jakobson.

It is sometimes argued that binary features are necessary for phonological theory. Ladefoged (1971) and Conteras (1969) present some convincing arguments against this position and point out that the Chomsky and Halle (1968) phonology is not without multivalued feature specifications (the stress rules). Even if phonological theory required the binary principle, it would not follow that the perceptual dimensions involved in phonemic recognition must be binary.

Jakobson Summarized

To summarize the argument thus far: The phonemic function of distinctive features is of primary importance in the Jakobsonian system. This function is relevant in that what people hear when they listen to speech is powerfully influenced by the phonemic pattern of their native language. There are three major constraints on feature postulation for Jakobson. Features must be binary, universal, and minimal in number. The binary principle is not well founded on either perceptual or linguistic grounds. The requirements of universality and minimal number are synergistically related. The fewer the number of features, the more powerful the universal linguistic implications of the theory appear to be, and the combination of the two constraints does in fact constrain a distinctive feature representation in a way that a language-specific or a numerically unlimited set of features would not. However, it is by no means clear that these constraints are appropriate, unless one is prepared to entertain extrinsic assumptions of the kind mentioned in Chapter I - that the human auditory apparatus is genetically constituted to perceive speech sounds in a unique and rather specific manner. It is not clear that this assumption is warranted or testable at the present time.

One further point should be made in reference to the Jakobson, Fant, & Halle (1951) system, though it applies to all current linguistic feature systems. For purposes of constructing a perceptual model, the requirement that every feature have a readily specifiable and measurable acoustic correlate is both too weak and too strong. It is too weak in the sense that there are an indefinite number of readily measurable acoustic attributes of auditory stimuli (that may be employed for stimulus subcategorization), which may or may not be reliably detectable by the auditory apparatus. It is too strong in the sense that, what is a simple and readily detectable auditory parameter to some biological sound analyser, may be a highly complex integration of those parameters that the acoustic engineer is able to characteriize within the limits of his instrumental technology. Neurophysiologists have become quite conscious of this embarrasing fact in recent years (Whitfield and Evans, 1965; Worden & Galambos, 1972, in passim).

Features in Generative Phonology

Generative phonology has given particular importance to the phonological function of distinctive features. There are numerous illustrations available which show how the choice of features can be governed by phonological considerations. A notable source of such illustrations lies in the proposals that have been offered regarding some perceived inadequacies of the Sound Pattern feature system. For example, several writers have pointed to the inadequacy that the Sound Pattern feature system "does not permit us to formally express the fact that lip based sounds (+anterior, -coronal) and round sounds (+round) form a natural class [Ladefoged & Venneman, 1971, p. 14].

For example, the sound change /w/→ /v/ is a widespread diachronic phenomenon. The naturalness condition (Chomsky & Halle, 1968, p. 335) stipulates that sounds which are prone to undergo such phonological alternations be economically representable by the feature system. Conversely, "unnatural" (rarely or never alternating) phone classes ought to be much more difficult to express by the feature notation. Similarly, it is quite common to find non-labial consonants assimilate to the labial point of articulation in the environment of a rounded vowel or glide (see Campbell, 1974 for several examples of "labial attraction" rules). Where this assimilation does not alter the primary point of articulation, but merely introduces a "secondary" articulatory overlay of lip rounding, the Chomsky and Halle feature system works. However, changes in primary point of articulation are not uncommon, and for these cases the feature system fails to express the change economically, and, more importantly, does not capture its assimilatory character. The evident solution to this problem is to postulate a feature (labial) which encompasses bilabial and labiodental closure in consonants and lip rounding in vowels and glides.

The question has been raised as to whether phonetic and phonological features are necessarily the same kind of entity. Ladefoged (1971a,b) has pointed out that among those features with the strongest phonological motivation are some that are not associated with any "single, measurable" acoustic or physiological property (such features as consonantality, labiality, or the stress feature). He proposes a categorical distinction between these features - referred to as "cover" features - and "primary" features that do have a measurable phonetic referent.

> Any empirical theory has to have a number of primitives which are definable in terms of concepts which belong outside the theory. In the case of phonological theory, these are prime features which are definable in terms of acoustic or physiological properties of sounds... In addition there are phonological features that are not themselves prime features but disjunctions of values of prime features; ...they are cover terms for certain values of related prime features [Ladefoged & Venneman, 1971, p. 13].

> ... the relationship between them [prime and cover features] is of the form indicated by feature redundancy rules. The number of prime features must, as in any theory, be minimal; but the number

> of derived features constructed from them must be sufficient so that we can give explanatory formulations of linguistic phenomena [1971, p.23].

One possible formal objection to the use of cover features is that they appear to be insufficiently constrained by considerations that must necessarily be met if phonological rules are to have any explanatory power. In other words, the criteria for cover features do not seem to prevent the establishment of ad hoc feature classes that may be very convenient for the formulation of phonological rules, but may at the same time be quite "unnatural" from the standpoint of phonetic similarity.

As both Fromkin (1970) and Ladefoged (1971) have argued, the "naturalness" of a particular feature or the explanatory basis for a given phonological rule cannot be established on formal grounds, but reduces to one of two classes of considerations: namely, considerations of perceptual similarity and contrast or considerations of production. For this reason, the phonetic basis of a distinctive feature system will be a mixture of auditory and articulatory considerations. "Some features will be more easily interpretable in one way, and others in the other [Ladefoged, 1971, p. 7].

It would appear then that cover as well as prime features should be justifiable on grounds of either production or perception. But does this entail that they must be reducible to "simple" scalar physical properties? It has already been argued that the acoustic correlates of an auditory-perceptual variable may be neither simple, nor - for technical reasons - accurately measurable, but nonetheless real. The same would appear to hold in the domain of production. The fact that articulatory based taxonomies for phonetic description have existed for centuries and that various systems show a good deal of correspondence should not be allowed to obscure the fact that very little is known about the relevant control and feedback parameters involved in speech production. Should an articulatory feature system which is optimal from the standpoint of a "universal phonetic theory" reflect the geometry of the vocal tract - some spatial representation of a set of ideal articulatory target points? If so, then from what is known about somatotopic representation in the central nervous system (Mountcastle, 1970) one would be lead to suspect that the subjective geometry of the vocal tract is related to its physical proportions in some highly complex, non-linear fashion. Would it perhaps be better to base an articulatory feature system upon an analysis of synergistically operating groups of muscles? What role should be given to tactile as opposed to proprioceptive or auditory feedback?

Ladefoged's "cover" versus "prime" feature distinction may be regarded as one way of resolving a long-standing controversy in phonological theory (Trubetzkoy, 1969), namely, whether phonologists need or ought to be concerned with the phonetic basis of speech in arriving at phonologically well-motivated sound classes (feature systems). In terms of the concerns of this paper, the question may be turned around to ask, whether there is some plausible basis for the supposition that features posited on the basis of their phonological function have some relevance to a perceptual model. It has often been observed that not all the sound regularities described in phonological rules are readily explicable in terms of operational constraints on the perceptual or speech production systems. The appeal to "ease of articulation", for example, has most likely some validity for common assimilatory processes observed in many languages. On the other hand, many of the "phonotactic constraints" characteristic of a particular language appear to be quite arbitrary. These sequential constraints on sound combination introduce redundancies into the speech signal which could well be utilized by a perceptual mechanism having access to them through some internalized representation of the sound pattern of the language. Part of the phonological function of distinctive features is to provide an economical means of distinguishing such classes of permissible from non-permissible sound sequences (Halle, 1962; Chomsky & Halle, 1965).

Insofar as distinctive features are necessary or useful for rule writing and insofar as the rules capture information which could be utilized by a perceptual device, to enhance speed and reliability of operation, particularly under "difficult" listening conditions, an indirect and *prima facie* case can be made for phonologically motivated features as components of a speech perception model.

Consider the following argument: Many of the sequential constraints on sound combination in English (or any other language) are contingent upon the presence of linguistic boundary markers of one kind or another: morphemic boundaries, word boundaries, phrase boundaries, etc. (for a detailed discussion of the current status of boundary symbols in phonology, see Stanley, 1974). Of particular interest, largely because of the prominant role that they have played in discussions of generative phonology, are the class of rules variously known as "morpheme structure" rules or conditions (Stanley, 1970), or as "lexical redundancy rules" (Chomsky & Halle, 1968). The major function of such rules in generative phonology is usually characterized as one of economizing on lexical representations, or its complement, of maximally exploiting regularities in the sound sequencing of the language. For example, very few of the possible pairwise combinations of consonantal phonemes constitute permissible morpheme-initial consonant clusters in English. Hence many of the phonological features in lexical items containing an initial consonant cluster are predictable by rule and need not be entered in the lexicon. The relevance of this kind of "economic" consideration for a performance model has already been questioned (Chapter I).

However, there is another way of conceiving the functional role of morpheme structure rules that seems to be more germane to the problem of speech recognition. The determination of morphological boundaries is one of the necessary analytical operations for any model of speech recognition that employs a word or morpheme lexicon. (Words and morphemes are not, of course, synonomous, but for purposes of the present discussion the distinction is not important.) Morphological boundaries must be at least tentatively determined in order to associate items in lexical storage with "portions" of the continuously changing incoming auditory signal. Just as boundary symbols are employed as part of the structural description of phonological rules to "predict" information about the phonological feature specification of lexical items in the language, so certain specific collocations of features may be used to establish linguistic boundary markers that might otherwise have no overt phonetic manifestation in the signal.

However, phonotactic constraints may be represented quite adequately for the purposes of lexical economizing or linguistic boundary-marker assignment without any recourse to a subphonemic, distinctive feature system. They may simply be stated in terms of collocational restrictions on unanalysed segments (systematic phonemes) and certain linguistic boundaries. Conceivably, they may even be statable in terms of a (phonological) syllabary.

Therefore, the argument for the plausibility of phonologically motivated features from the possible perceptual utility of the rules that they permit the phonologist to write, is not a compelling, or even a strong one. And, as the phonological motivation for a feature is the only criterion for its postulation which is considered by the proposed "evaluation metric" of generative phonology, there is no reason to expect that the optimal feature set from the viewpoint of a generative phonology will also be optimal from a perceptual point of view.

CHAPTER III

PERCEPTUAL DIMENSIONS IN PHONEMIC RECOGNITION

Even a superficial contact with the literature reveals that a broad range of methodologies and experimental rationales have been used in the attempt to isolate and describe the major perceptual factors involved in phonemic recognition. In the interests of a coherent presentation of the substantive findings it is advisable to briefly indicate each of these methodologies, stating their possible interrelationships and respective limitations. The review of the published findings in this chapter will focus upon the work with consonantal sounds and emphasise the rationale for selecting a particular data base and method of data collection. Some of the most interesting work has been done with vowel perception (Pols, van der Kamp, & Plomp, 1969; Terbeek & Harshman, 1971; Harshman, 1971). This work is primarily of methodological interest to the present study. It will therefore be mentioned in the context of the evaluation of Multidimensional Scaling (MDS) and related data reduction techniques in Chapter IV.

Researchers differ considerably in their reliance upon phonological theory to provide the analytical framework for their studies. An important distinction can be drawn between those studies that rely upon some a *priori* feature scheme and those that derive features a *posteriori* from the data base. Most of the earlier studies relied upon an a priori feature system:

> ... they could explain perceptual patterns only in terms of the predetermined attributes without having the option of systematically exploring the possibility of the perceiver's utilization of more appropriate attributes [Singh, 1974, p. 56].

The fundamental problem for researchers who derive perceptual features a posteriori from the data is how appropriate (well motivated) are the theoretical presuppositions upon which the data reduction procedure is based?

Perceptual confusions have provided the most popular data base for isolating perceptual features in phonemic recognition. This approach rests on the assumption that sounds which are similarly valued on whatever perceptual attributes the listener uses in their recognition or discrimination will (other things being equal) show a higher probability of mutual confusion - particularly under difficult listening conditions of one kind or another. A confusion matrix in which the off-diagonal elements contain the relative frequency of misidentifications of each perceptual target with every other in the set should therefore contain (though not necessarily in readily accessible form) information about the set of underlying perceptual features involved. Because of the high reliability of perception under normal listening conditions, to generate analysable patterns of confusion scores in the off-diagonal elements of the matrix, it is usually necessary to subject the perceptual system to some kind of operational stress - via signal masking, distortion, or attenuation, or by arranging the task so that some performative limitation of the subject is exceeded to the point where he makes frequent errors. Miller and Nicely (1955), in their classical study of consonantal perception under white noise masking, and low and high-pass filtering, were the first to explore this approach. Their extensive data base has been used by numerous subsequent investigators (Wilson, 1963; Johnson, 1967; Cocoran et al., 1968; Wish, 1970; Hollaway, 1971; Shepard, 1972; Smith, 1973).

An attractive feature of confusion matrices obtained under conditions of signal masking is that the experimenter can be reasonably assured the data will be uncontaminated by extra-perceptual factors. However, when the subjects' performative capabilities are placed under stress - for example, by an interpolated memory task (Wickelgren, 1966) - the experimenter no longer has this assurance. This objection has also been raised (Shepard, 1972) against another common method of measuring perceptual proximity - subjective estimates of perceptual similarity.

Either direct or indirect similarity scaling may be used to generate a perceptual proximity matrix. Direct estimates of perceptual similarity may be obtained by a variety of psychometric procedures. Subjects may be required to make pairwise ratings of the stimuli on a categorical or continuous scale of overall perceptual similarity. Alternatively, their experimental task may be to simply choose between two alternative responses, the one which is "most like" a given stimulus as standard.

A measure of perceptual similarity may also be generated indirectly from ratings on a number of semantic scales rather than a single global similarity scale. The set of phoneme targets may be rated by subjects with respect to a number of verbal-descriptive scales, thought to capture various perceptually relevant qualities of sounds of speech. The degree of correlation between the profiles of scores on the set of semantic scales for any two phonemes may be taken as an index of the similarity of the two phonemes. Of course, the accurracy of any such indirect estimate of similarity is predicated upon an appropriate choice of descriptive scales and relative scale weightings.

If the experiments are properly carried out, the results of direct and indirect similarity scaling should be highly correlated. Similarity scaling has the advantage over confusability indices of providing perceptual proximities under listening conditions that are free of gross signal distortion or some other abnormal, error-inducing factor. But, on the other hand, similarity scaling is based upon what is, for the listener, a rather artificial task.

Discriminatory reaction time (DRT) has recently been explored as a measure of perceptual proximity (Weiner & Singh, 1974), based upon the assumption that closely related perceptual targets will require longer discrimination latencies. The technique looks promising.

Perceptual Confusion Matrices

Miller and Nicely (1955) used an *a priori* feature system to evaluate the effects of 17 experimental conditions of white noise masking and signal filtering upon the identification of 16 consonants embedded in a /Ca/ syllabic frame (see Table 3.1 for features, conditions, and stimuli).

TABLE 3.1

DETAILS OF THE MILLER AND NICELY (1955) EXPERIMENT

Paradigm:	Confusion matrices collected under conditions of (a) white-noise masking; 6 S/N ratios ranging from -18dB to +12dB, fr. response 200-6500Hz (b) low-pass filtering; 6 conditions, 200-300Hz, 200-400Hz, 200-600Hz, 200-1200Hz, 200-5000Hz (c) high-pass filtering; 6 conditions, 1-5kHz, 2-5kHz, 2.5-5kHz, 3-5kHz, 4.5-5kHz
Subjects:	5 female subjects who served as talkers and listeners
Stimuli:	/p,t,k,f,θ,s,š,b,d,g,v,ð,z,ž,m,n/ in a /Ca/ frame

The central part of their data analysis was concerned with the relative amounts of information lost for each feature in transmission under different listening conditions. For any feature, the amount of information transmitted was (not surprisingly) inversely monotonically related to the severity of the white noise masking or the degree of filtering. What is of interest however, is the relative imperviance of some features compared with others, and how the relative imperviance of the features to transmission loss is differentially affected by the three basic conditions of white noise masking, high, and low-pass filtering.

Under all but the most unfavourable condition of white noise masking (-18dB, where the subjects' responses are virtually random) the features of nasality and voicing were better preserved than duration, affrication, and place of articulation. White noise primarily masks the auditory cues carried by the components with lower intensity, which for most speech sounds are in the higher frequency regions of the acoustic energy spectrum. Its effect on intelligibility is similar to that of low-pass filtering (as the Miller and Nicely data show).

Miller and Nicely's analytical techniques were inadequate for the problem of deriving an optimal set of features that would account for the pattern of misidentifications among the off-diagonal elements of the confusion matrix. Wilson (1963) applied Torgerson's (1958)

MDS procedure to the -12dB S/N condition of the Miller and Nicely data. This particular level of white noise masking was chosen for being:

> ... low enough to permit a considerable amount of differentiation between the consonants yet high enough to permit application of the analytical techniques... [Wilson, 1963, p. 89].

He derived two sets of interpoint distances from the confusion matrix and (following Torgerson) reduced these two 16 dimensional configurations in both cases to four dimensions by the Principle Axes method of factor analysis. Wilson used two quite different formulae for deriving interpoint distances from the frequencies in the confusion matrix because of the unsolved problem of choosing an appropriate function to relate raw data scores to distances. This problem is particularly accute with the older "metric" forms of MDS.

The two (orthogonal, unrotated) factor solutions agreed with one another and the Miller and Nicely findings to an interesting degree. For both MDS solutions (one based on Shepard's (1957; 1958) distance formula, the other on his own derivation) the first factor clearly differentiated all the voiced from the voiceless consonants. The nasals were clearly distinguished from the other consonants by the second factor loadings for both solutions. However, both solutions do not clearly indicate a simple nasality factor:

> For Shepard's measure, Factor II gives sizable positive loadings to the longer duration consonants, /s/, /s/, /z/, /z/ and to /d/ and /g/ and sizable negative loadings to the two consonants /m/ and /n/. Obviously, this is a complex and not easily interpreted dimension. For Wilson's measure, Factor II gives sizable positive loadings to the two nasal consonants and near zero loadings to the others and so is obviously a nasality factor [1963, p. 93].

There was no clearly discernable agreement between the two solutions for factors III and IV. Nor did these dimensions appear to be readily interpretable. (But Wilson did not attempt to rotate the factor axes in order to improve the interpretability or maximize the agreement between his two solutions.)

Shepard (1972) reanalysed the Miller and Nicely data using a "non-metric" MDS technique (Kruskal, 1964a,b) and Johnson's (1967) method of hierachical clustering. By pooling the six confusion matrices elicited under white noise masking, he obtained an optimal two-dimensional solution, accounting for 99% of the variance (see Figure 3.1 below). Shepard's MDS solution shows essential agreement with Wilson's findings and further indicates the prominence of the features of voicing and nasality under white noise masking. However, while Shepard's results indicate that a considerable dimensional reduction of the confusion matrix is achievable, the distribution of sounds along the reference axes does not unambiguously favour a simple, orthogonal, "two feature" interpretation. (Consider the

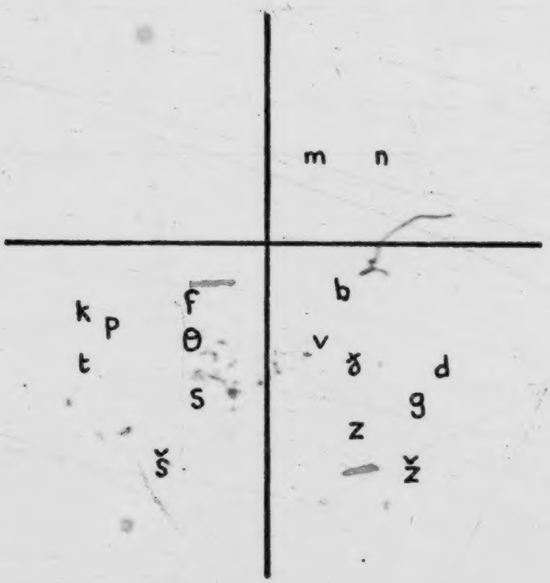


Fig. 3.1 Reanalysis of Miller & Nicely (1955) data. From Shepard (1972).

second dimension. It is intuitively obvious that /b/ and /f/ are less "nasal" than /m/ or /n/ but it is by no means clear that this same quality even better differentiates /f/ and /b/ from /s/ and /z/ as a straightforward interpretation of the configuration would imply.)

Superimposed on the configuration (3.1) are the 3 and 5 group levels that the hierachical clustering analysis suggested were the most reliable subgroupings of the consonants. Beyond the two features of voicing and nasality, the hierachical clustering solution is not predictable from Miller and Nicely's a priori feature set.

Detailed analysis of the effect of masking level showed that while varying the level of white noise masking predictably affected the overall level of confusion, the recurrence of the same clusterings at each S/N ratio indicated that:

> The internal pattern of confusions was essentially invariant. Indeed with respect to the spatial representation, the effect of adding a given amount of white noise seemed to be almost entirely confined to a reduction of all interpoint distances by the same constant factor [Shepard, 1972, p. 109].

With respect to the other listening conditions:

> Generally the pattern resulting from low-pass filtering is remarkably like the pattern resulting from the addition of broadband noise... the only notable difference seems to be that /f/ and /θ/ group with the unvoiced stops /ptk/ in the "flat" conditions but with the other unvoiced fricatives /ss/ in the low pass conditions [1972, p. 101].

However, the pattern that emerged through cluster analysis of the confusions under the high-pass conditions differed "radically" from that obtained under white noise masking and low-pass filtering.

It may be useful to conceptualize the effect of different kinds of filtering and masking as differentially affecting the weightings of a set of underlying features - enhancing the relative prominence of some features under one condition and supressing or erasing other features normally used in phonemic recognition. Carrol and Chang's (1970) INDSCAL method of MDS makes just this kind of assumption. INDSCAL, a three mode scaling procedure, derives a "group" configuration on the assumption that the same dimensions are operative in each of the "individual" matrices that comprise the third mode, though the relative weightings of the dimensions in determining the interpoint distances may vary across individuals. Wish (1972) applied INDSCAL analysis to all 17 of the Miller and Nicely matrices. In addition to the dimensions of Nasality and Voicing (Dimensions I and II respectively), several additional factors were extracted namely: "voiceless stop vs. voiceless fricative" (duration?); "second formant transition"; "sibilance"; and "sibilant discrimination".

Questions Posed by Miller and Nicely's Data

All reported analyses of Miller and Nicely's data concur in attributing perceptual prominence to the Voicing and Nasality features. Two questions however arise that bear on the generalizability of these findings. (a) To what extent does the preponderance of high frequency attenuation in Miller and Nicely's experiment over-enhance the effect of low frequency signal components and thus yield a distorted picture of perception under normal listening conditions? (b) What impact does Miller and Nicely's particular choice of consonants have on the results of the experiment?

Specifically, the lack of resonant sounds in the data set raises the question of whether the prominence of "nasality" is not at least in part attributable to a more general auditory feature separating not only the nasals, but also the glides and liquids from sounds characterised by a turbulent noise source.

Wang and Bilger (1973) have recently reported a study of perceptual confusions under white noise masking and (flat frequency) signal attenuation which permits a partial answer to (a) and (b) above. They examined consonantal confusions with four sets of stimuli under (i) different S/N ratios of white noise masking and (ii) different signal levels under quiet listening conditions (see Table 3.2 for details).

TABLE 3.2

DETAILS OF THE WANG & BILGER (1973) EXPERIMENT

Paradigm: Confusion matrices collected under conditions of:
(a) white noise masking; 6 S/N ratios ranging from -10dB to +15dB
(b) signal attenuation without noise masking ("quiet" condition)

Subjects: 16 paid volunteers assigned to 4 listening groups,

	Consonants (vowel = /a/)
CV-1	/p,t,k,b,g,d,f,θ,s,š,v,ð,z,ž,č,j/
VC-1	/p,t,k,b,g,d,f,θ,s,š,v,ð,z,ž,č,j/
CV-2	/p,b,c,j,l,r,f,s,v,z,m,n,h,hʷ,w,y/
VC-2	/p,b,g,ŋ,m,n,f,θ,s,š,v,ð,z,ž,č,j/

Wang and Bilger also employed an a priori set of features to analyse subjects' perceptual confusions. However their analytical method was more powerful than that of Miller and Nicely. In addition to simply comparing the information transmission levels for different features in order to distinguish those that are perceptually prominant (stable) from those that are weak (subject to transmission loss), Wang and Bilger were concerned with minimizing the internal redundancy of the feature set as a whole. To this end they developed:

> ... a sequential method of analysing transmitted information which systematically identifies from among a number of features, those on which performance is high, and which takes the internal redundancy of the features into account in doing so. This is accomplished by partialling out, in each iteration, the effects of features identified in earlier iterations. The analysis also allows us to determine what proportion of the total transmitted information is accounted for by the features identified as perceptually important. The procedure and the rationale behind it may be loosely interpreted as the information analogue of a stepwise multiple regression analysis [Wang & Bilger, 1973, p. 1249].

They were thus able to input a large and highly redundant set of features into the analysis without risk of loosing those features that make a significant independent contribution to the identification of the consonants, as distinct from those whose reliability of transmission may be accounted for by their covariation with one or another more reliably transmitted feature. The feature set used by Wang and Bilger was simply a combination of those of Chomsky and Halle (1968), Singh and Black (1968), Wickelgren (1966), and Miller and Nicely (1955).

Table 3.3 lists, in rank order, the features that emerged as significant for the different stimulus sets under conditions of white noise masking (averaged over all S/N ratios). Column one of Table 3.3 gives the results of applying Wang and Bilger's algorithm to Miller and Nicely's pooled conditions of white noise masking. The prominence of nasality and voicing is clearly apparent for all sets of

TABLE 3.3

PERCEPTUAL SALIENCE OF FEATURES - WHITE-NOISE CONDITION
(WANG & BILGER, 1973)

	Stimulus Set				
	Wang & Bilger				Miller & Nicely
rank order ↓	CV-1	VC-1	CV-2	VC-2	
	Voicing	Voicing	Nasal	Nasal	Nasal
	High	Coronal	Vocalic	Voicing	Voicing
	Sibilant	Continuant	Round	Frication	Duration
	Frication	Place	High	Open	Continuant
	Open	Place(W)	Voicing	Back	Place(SB)
	Place(W)		Sibilant	Sibilant	
			Coronal	Place(SB)	
			Place(W)		
			Open		

All features that made a significant (independent) contribution in the Sequential Information Analysis (see above) are listed in rank order.

data where these features are applicable. (No nasals are included in the CV-1 or VC-1 sets.) Unfortunately, the question raised earlier with respect to a more general resonance feature remains unanswered because not one of the 19 input features serves to group the resonants together. It is notable however, that the only stimulus set containing a mixture of nasals and other resonants (set CV-2) yields nasal, vocalic, and round as its three strongest features (i.e., the subsets /m,n/; /l,r/; /h,w/).

Other features that emerged prominantly under white noise masking were: Open (for a definition of this and any other feature mentioned elswhere in this paper see the alphabetic listing in Appendix A), Sibilance, and Continuance.

Returning to the question of the perceptual ove[illegible] enhancement of low frequency signal components under white noise masking, Wang and Bilger observed that:

> Voicing and nasality are well perceived in the presence of masking, but their intelligibility drops relative to that of other features in quiet [Wang & Bilger, 1973, p. 1254].

Singh (1971) observed the same effect for the perception of minimal pair phonemic differences under noisy vs. quiet conditions.

Perceptual Confusions under Non-Noisy Listening Conditions

Singh and Black (1966) obtained perceptual confusion matrices on 22 intervocalic consonants as spoken and heard by native speakers of four different language groups (Hindi, English, Arabic, Japanese). An a priori set of seven features (Voicing, Nasality, Aspiration, Frication, Place, Duration) was used in an information transmission analysis identical to that of Miller and Nicely.

> The single striking outcome of the present study lies in the rank orders of the seven channels in the relative amounts of information per channel - that is in the "importance" of the channels. A single rank order in this regard obtains for all the listening groups: (1) nasality (2) place (3) liquid (4) voicing (5) duration (6) frication and (7) aspiration [Singh and Black, 1966, p. 387].

However, the application of Wang and Bilger's expanded feature set and analytical algorithm (which takes account of the internal redundancies of the features) to the Singh and Black data suggest a rather different ranking of features in terms of perceptual prominence (see Column 1 of Table 3.4). Most notably, controlling for internal redundancy, lowers the salience of the Place feature and raises that of Frication.

Graham and House (1971) studied errors in consonantal

TABLE 3.4

PERCEPTUAL SALIENCE OF FEATURES - "QUIET" CONDITION
(WANG & BILGER, 1973)

Wang & Bilger				Singh & Black	Graham & House
CV-1	VC-1	CV-2	VC-2		
High	Sibilant	Round	Nasal	Nasal	Sibilant
Voicing	Duration	Vocalic	High	Vocalic	Duration
Back	High	Nasal	Back	Frication	Frication
Sibilant	Voicing	Sibilant	Duration	Back	Voicing
Frication	Continuant	Duration	Voicing	Voicing	Anterior
Duration	Place(W)	Anterior	Frication	Place(SB)	Round
Place(W)		Coronal	Place(W)	Open	Conson.
		Voicing		Place(W)	Nasal
		Open			Open
					Place(W)

discrimination made by young children (aged 4 years) who were asked to give "same or different" judgements to pairs of orally presented CV syllables. The 16 consonants used by Graham and House are given in Figure 3.2, which shows the two-dimensional configuration obtained by Singh, Woods, and Tishman's (1971) MDS reanalysis of the original confusion matrix. Dimension I of Figure 3.2 appears to be a temporal factor separating the stops from the continuant sounds. Dimension II, on the other hand, clearly differentiates the sibilants from the resonant consonants.

The MDS configuration for the children's perceptual errors under non-noisy listening conditions is quite different from that obtained (by the same scaling technique) from adults' phonemic misidentifications elicited under white noise masking (Figure 3.1). The difference between these two scaling solutions may be tentatively accounted for

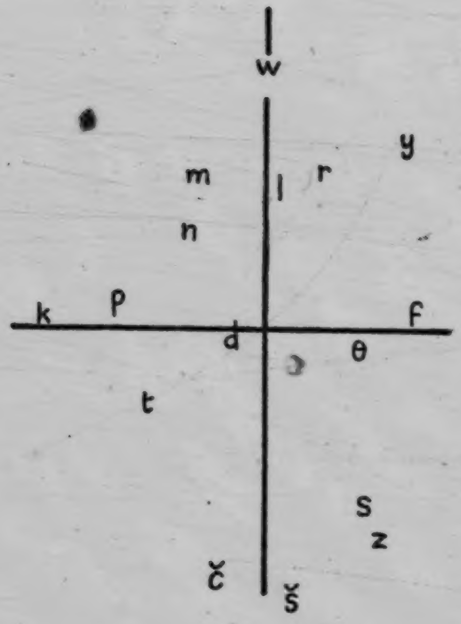


Fig. 3.2 Reanalysis of Graham and House (1971) data
Phonemic Confusions of Young Children
from Singh and Woods, 1971.

by (1) a rise in the perceptual salience of the durational characteristics of the speech signal under non-noisy listening conditions, together with (2) a reassertion of the turbulent noise characteristics which would be most adversly affected by white-noise masking, and (3) a relative diminution of the prominence of the low frequency signal components in the absence of high frequency masking noise.

In short, discrepencies between the two configurations may be attributable to the differences in listening conditions rather than the obvious task and subject differences. This is a rather bold interpretation, but it is supported by internal comparisons within the Wang and Bilger data. Compare the salience of the Duration feature for the four Wang and Bilger stimulus sets in Table 3.3 (white-noise condition) with Table 3.4 (quiet condition).

Duration appears as a significant feature in all four of the stimulus sets in the "quiet" condition where errors were induced by lowering the signal level. However it never appears as a significant perceptual feature for these same sets of stimuli under the noisy listening condition in which errors were induced by lowering the S/N ratio. The same effect may be demonstrated by following the changes in rank status of the duration feature across the six increasing S/N ratios (see Table 3.5). In the case of the Voicing feature, the opposite trend is apparent:

TABLE 3.5

PROMINENCE OF DURATION AS A FUNCTION OF S/N RATIO (WANG & BILGER, 1973)

Syllable set	-10dB	-5dB	0dB	5dB	10dB	15dB
CV-1	ns	ns	ns	ns	7	5
VC-1	ns	ns	ns	5	2	1
CV-2	ns	ns	ns	ns	ns	5
VC-2	ns	ns	ns	ns	ns	3

n.b. Numerals represent rank status of Duration feature among others that emerged as significant in the Sequential Information Analysis. The characters ns indicate when the feature failed to make a significant contribution under a given level of white-noise masking.

However, not all the available data from confusion matrices support the hypothesis that differences in

TABLE 3.6

PROMINENCE OF VOICING AS A FUNCTION OF S/N RATIO (WANG & BILGER, 1973)

Syllable set	-10dB	-5dB	0dB	5dB	10dB	15dB
CV-1	1	1	1	1	1	2
VC-1	1	1	1	1	1	2
CV-2	1	1	4	5	6	8
VC-2	2	2	2	2	2	4

listening conditions alone may account for differences in derived perceptual configurations and feature weightings. We would expect, for example, that duration should have emerged as a significant feature in Wang and Bilger's reanalysis of Singh and Black's cross-language confusion data (Table 3.4 above).

Perceptual Proximities via Similarity Scaling

Peters (1963) appears to have done the first experimental study of perceptual features in phonemic recognition utilizing MDS of subjective similarity judgements. Peters employed two sets of stimuli - one comprising 28 consonants in a /Ca/ frame and the other identical with the Miller & Nicely set. Procedures differed slightly for obtaining similarity ratings on the two sets of stimuli. On the set of 16 consonants, subjects made pairwise ratings of the syllables on a 9 point scale of overall perceptual similarity, shortly after having spoken the syllables aloud. The subjects' raw score ratings were entered in a 16 x 16 similarity matrix and treated as absolute distances for input into Torgerson's (1958) scaling procedure. Only three subjects judged the 28 stimulus set. Nine were used for the 16 syllable set and separate scaling solutions were obtained for each subject. The obtained solutions ranged in dimensionality from 2 to 5.

> Examination of the data indicated that two or three dimensions were relevant and the higher dimensions, when they appeared, did not seem to be necessary for adequate description of the data [Peters, 1963, p. 1987].

Peters, "in keeping with previous work," interpreted his results in articualatory terms:

> The results indicate that manner, voicing, and place of articulation, are of importance in this respective order... [1963, p. 1988].
>
> The major grouping of the consonants was by manner, with either place or voicing represented as a within group dimension [1963, p. 1987].

The reliability of the individual configurations is rather questionable but the accuracy of Peters' observations is supported by Shepard's (1972) reanalysis of the pooled proximity matrix derived from the nine subjects who judged the 16 consonants. Hierachical clustering analysis yielded stable clusters at the 4 group and the 8 group level (see Figure 3.3).

The striking feature about these results, in contrast

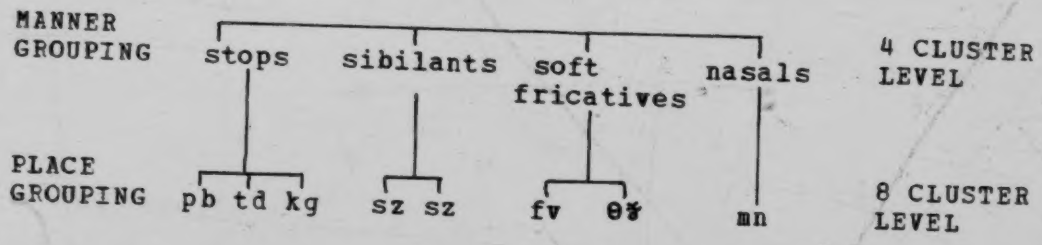


Fig. 3.3 Hierachical Clustering Analysis (Shepard, 1972) of Peter's (1963) pooled similarity matrix

to the clusterings yielded by the Miller and Nicely data and other confusion matrices (Singh and Black, 1966; Wang and Bilger, 1973) is the weakness of the voicing dimension. Unfortunately, Shepard did not subject Peters' pooled similarity matrix to an independent MDS analysis, but simply embedded the above clusters in the two-dimensional space of the Miller and Nicely data. (And Peters does not supply the raw data matrices for individual subjects.) It is difficult, therefore, to form a clear idea of the basis that the subjects may have used for their judgements which lead to the formation of clusters along the lines of traditional manner of articulation categories. Inspection of the individual scaling solutions in two dimensions, which Peters does present, shows that in the majority of cases, one axis polarizes the stops and the fricatives and the other orthogonal axis tends to separate the nasals from the other consonants. The configuration underlying the clustering would seem to be not unlike that of the Singh et al. (1973) reanalysis of the Graham and House (1971) data (Figure 3.2). Once again, it is tempting to attribute the apparent weakness of Voicing to the absence of high frequency masking. However, Shepard (1972) suggests that other factors may be operative:

> One hypothesis that might explain this apparent suppression of the usually salient feature of voicing... is that Peters' subjects treated this task as an analogy task rather than a pure similarity task
> [Shepard, 1972, p. 107].

Analogical reasoning, in which the subject systematically disregards otherwise prominant perceptual qualities of the stimuli, could have taken place. But this hypothesis raises the problem of explaining why subjects would consistently choose to disregard this particular perceptual feature. If analogical reasoning, essentially independent of perceptual processes, were affecting the subjects' performance, it seems more reasonalble to expect that this would simply introduce further heterogeneity or "noise" into the data. A third hypothesis (perhaps linked with the second) is that subjects relied significantly upon articulatory cues in making their similarity judgements; cues which are not employed in an identification task. Indeed, Peters' procedure favoured the use of articulatory cues, and one would expect Voicing to be a relatively "unmarked" feature in terms of tactile or proprioceptive feedback. Further data is obviously needed to resolve these questions.

Black (1968) collected estimates of subjective similarity between pairwise combinations of 24 consonants embedded in a /CV/ frame. Five vowels were used and 24 speakers each recorded a subset of the stimuli. On any given trial the pair of stimuli differed only with respect to the consonant and not the vowel or the speaker. (No motivation was given for this peculiarity in the experimental design.) A group similarity matrix was obtained and subjected to factor analysis (with varimax rotation). Black extracted no less than 12 factors. Factor I emerged as bipolar and he interpreted it as "sonority or a smooth-rough dichotomy." Factor II showed heavy positive loadings on the glides but also had a significant negative loading on the affricate /č/. Factor III was monopolar, showing significant positive loadings exclusively on the "soft" fricatives /f,θ,ð,v/. Factor IV appeared to be a sibilant dimension (hard frication). Factors V to VII separated out the pairs of stops /k,g/, /t,d/, and /p,b/ respectively, from the other consonants. The loadings were too few and too weak to justify interpretation of any of the other factors.

Because of questionable technical procedures that Black employed, his data were reanalysed for this study (with a principal components analysis and varimax rotation). The pattern of eigenvalues obtained from the principal components analysis indicated that no more than seven dimensions were justified by the data. Factor I clearly opposed the "hard" fricatives to the resonants /h,l,r,hw, and m/. Factor II also appeared to be a hiss - -resonant dimension, but this time opposing the "soft" fricatives /θ,ð,f,v/ to the resonants /l,y/. Factor III was bipolar and difficult to interpret as a single dimension with strong positive loadings on /t,d/ and negative loadings on /w,hw,r/. Factor IV separated the sibilants /s,z, and, though less strongly, š/ from the other consonants. Factor V clearly differentiated the two nasals. Factors VI and VII separated out the pairs of stops /p,b/ and /k,g/ respectively (see Appendix B for further details of the analysis).

To gain an overview of the perceptual relationships between the consonantal targets, it is useful to graph the first two principle component factors, which together account for 56.08% of the common factor variance and 47.69% of the total variance (see Figure 3.4). The graphical representation of the perceptual relationships between the consonants (3.5 above), and the pattern of factor loadings, both in the original study and the reanalysis, show a high degree of general agreement between Black's (1968) data and that of Peter's (1963), though such agreement could easily be obscured by superficial differences in analytical technique.

The plot of the first two principle components shows the major grouping by "manner of articulation" that Peters observed and Shepard (1972) later corroborated with clustering analysis. The major "resonant" group is consistent with Peter's "nasal" cluster - /m/ and /n/ being the only representatives of this group included in the 16

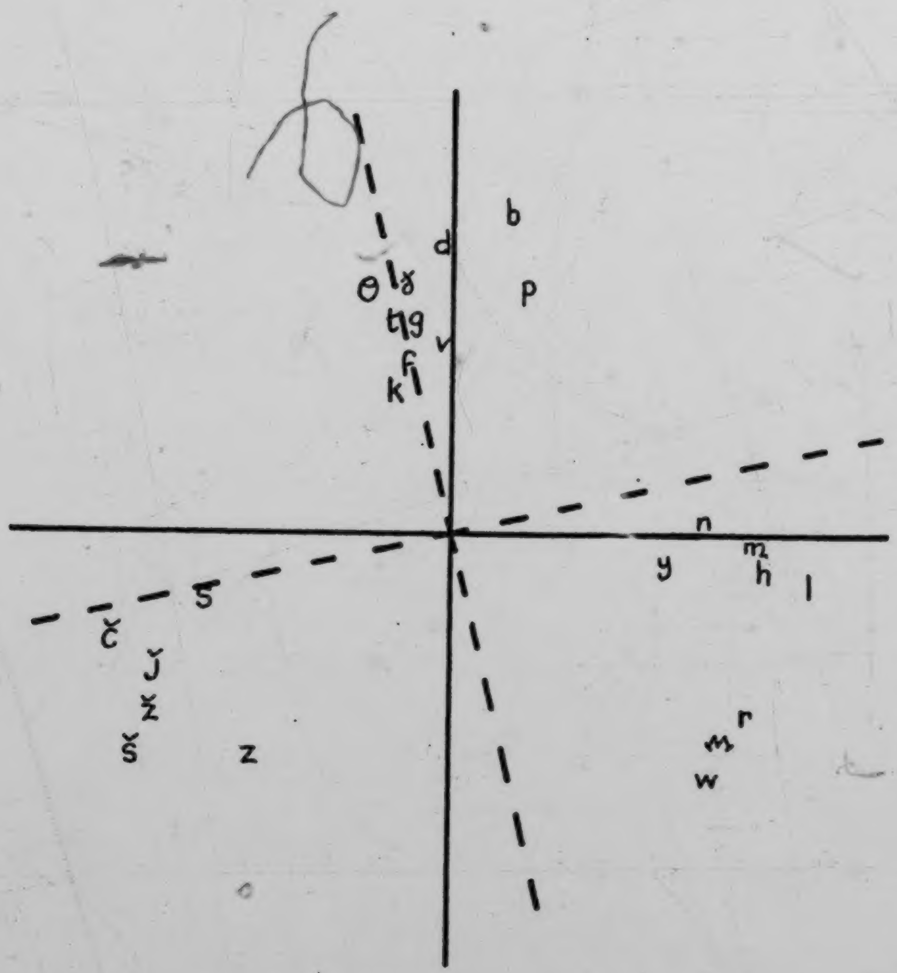


Fig. 3.4 Reanalysis of Black's (1968) Data
Similarity Rating of Consonantal phonemes
Principal Components analysis

Miller and Nicely consonants. The "stops" would be separated from the "soft fricatives", but for the fact that the third component is not shown in the graph (3.4), thus yielding a basic four-cluster configuration of : stops, sibilants, soft fricatives, and resonants. Varimax factors V to VII in Black's analysis and factors III, VI, and VIII in the reanalysis further corroborate the weakness of the voicing feature in the scaling of similarity judgements under non-noisy listening conditions.

Pruzansky (1971) used a computer-controlled sorting apparatus to initiate the presentation of 16 /Ca/ syllables (the Miller and Nicely set). Subjects located the sounds with respect to one another by arranging 16 pegs (one for each syllable) on a 16 x 16 pegboard. The Euclidean distances between pegs in the subject's final configuration on the board served as input to the Carroll - Chang (1970) INDSCAL MDS program. The author reports:

> The resulting group stimulus space revealed a separation between continuants and stops. Nasals were clustered. Pairs of stops differing only in voicing were grouped together [Pruzansky, 1970, p. 85].

Jeter and Singh (1972) studied perceptual similarity judgements of eight English consonants /b,t,d,f,v,s,z/ separately presented under either the auditory or the visual mode ("phonemic" vs. "graphemic" similarity). The ABX trial method was used to obtain auditory similarity ratings. The two group (n=60) similarity matrices (the phonemic and the graphemic) were scaled by Kruskal's method of MDS. The "stress" rating - Kruskal's index of goodness of fit between the input proximities and the derived distances - for the auditory mode matrix which is of primary interest here, was high: too high in fact to generate confidence in the reliability of the final solution. Dimension I of the derived three-dimensional configuration:

> ... could be clearly identified as a manner feature: all stops were separate from all continuants [Jeter & Singh, 1972, p. 705].

Dimensions II and III resisted clear interpretation. Multiple regression analysis showed that a three binary feature system of: place (labial - non-labial), manner (stop - continuant), and voicing best predicted - in that order - the derived interpoint distances for the auditory mode set (R=0.677). The resonant dimension is notably absent from the stimulus set.

Singh, Woods, and Becker (1972) reported a much larger study of perceptual similarity scaling, using 22 consonants /p,b,t,d,c,j,k,g,f,v,θ,ð,s,z,h,š,w,r,l,y,m,n/ and three similarity scaling methods (equal-appearing interval scaling (SF), magnitude estimation (ME), and triadic judgements (ABX)). A group matrix was accumulated for each of the scaling methods. The three group matrices were analysed separately by Kruskal's (1964) method and compositely by Carroll & Chang's INDSCAL method. Five dimensions were extracted from the INDSCAL analysis (see Table 3.6). Substantial differences between the derived perceptual configurations under the three scaling methods were noted. These differences are partly characterised by the relative weightings of the INDSCAL features for each of the group matrices (Table 3.6). There seems to be some question about the uniqueness and stability of this solution:

TABLE 3.7

INDSCAL DIMENSIONS AND WEIGHTINGS
FOR THREE SCALING METHODS
SINGH, WOODS, AND BECKER, 1972.

	DATA COLLECTION WEIGHTS		
	SF	ME	ABX
SIBILANT - NON-SIBILANT	0.349	0.380	0.562
FRONT - BACK	0.351	0.401	0.291
PLOSIVE - NON-PLOSIVE	0.277	0.260	0.310
VOICELESS - VOICED	0.306	0.277	0.217
NASAL - NON-NASAL	0.315	0.248	0.174

> The scaling was repeated in five dimensions with several different starting configurations. The clearest interpretation was found for the five dimensional space whose interpoint distances correlated at 0.78 with the data [Singh et al., 1972, p. 1709].

Differences between the three scaling methods are best observed when the three group matrices are scaled separately. Singh et al.'s criterion for determining the optimal dimensionality for the MDS analysis of the three group matrices is questionable. For all three matrices, Kruskal's stress criterion, and the plot of the tau correlation between derived distances and input similarities, would seem to indicate a two-dimensional solution (see Singh et al., 1972, p. 1704) and not the three and four dimensional solutions that the authors chose. Consequently, Singh et al.'s data were reanalysed with a Kruskal MDS routine (Euclidean distance metric). The resulting two-dimensional configurations for the three sets of data are given in Figure 3.6 below.

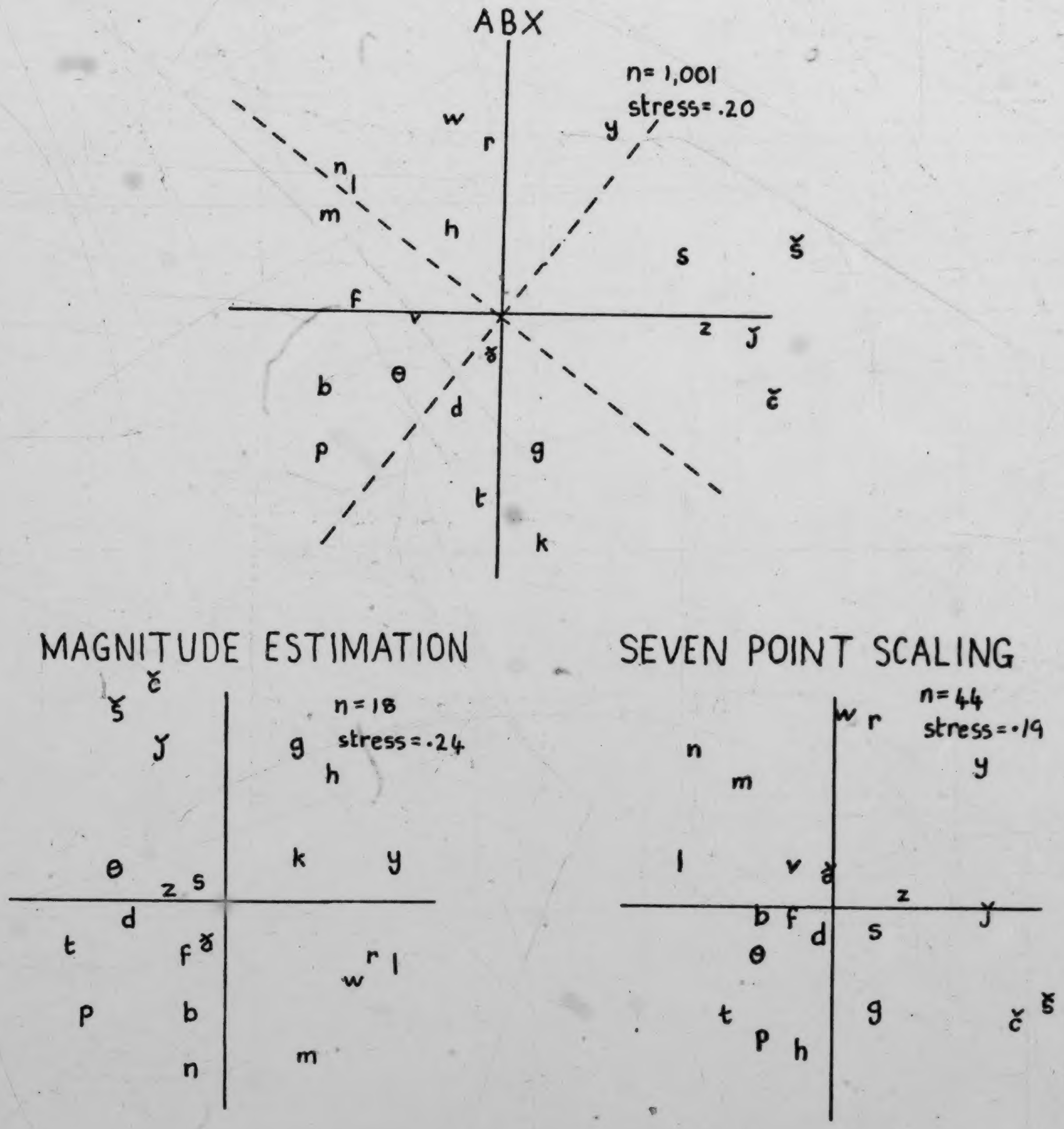


__Fig. 3.5__ Reanalysis of Singh, Woods, and Becker (1972)
Three Similarity Rating Methods
- Kruskal Scaling Two-dimensional Solution

Conclusions

The development of methods for establishing perceptual features involved in phonemic recognition, a posteriori, on the basis of a set of experimentally generated proximities, has freed the researcher from a certain amount of conceptual tyranny exercised by traditional phonetic taxonomies. However, it has by no means supplanted commonly used phonetic categories.

None of the traditional a priori features can claim ubiquitous perceptual prominence across all methods that have been used to estimate perceptual proximities, though some, clearly, are generally more important than others. Nasality is a particularly strong feature, though much of its prominence may be attributable to a more general Resonance factor with which it is confounded in the set of 16 Miller and Nicely phonemes employed in many of the published studies. Voicing emerges as a strong perceptual feature under white-noise masking. There is some disagreement about its status under quiet listening conditions. The Sibilants show a strong tendency to cluster together under non-noisy listening conditions. There is virtually no support for a feature such as Stridency which groups both the "strong" and the "weak" fricatives. Place of Articulation, which has always been problematical from the viewpoint of phonetic description is also unclear perceptually. The Labial-Nonlabial and the High-Nonhigh (palatovelar vs. non-palatovelar) contrasts appear to have significant status, at least in proximity measures based upon similarity judgements. There is no apparent justification for grouping labials and velars (Jakobson's grave feature) on the basis of similarity ratings or confusion matrices - the familiar acoustic justification notwithstanding.

The Stop-Continuant or Duration feature appears to be strong under quiet, but weak under noisy, conditions. It appears to be one of the factors primarily responsible for the observation that under good listening conditions, the consonants tend to cluster in accordance with traditional Manner of Articulation categories.

Many, but by no means all, the apparent discrepencies in the reported studies of perceptual relationships among the consonantal phonemes, can be readily attributed to peripheral signal masking effects, or to the composition of the experimental stimulus set. In particular, more information is needed about the impact of the choice of scaling technique upon the stability of the derived perceptual configuration.

The independence from extra-perceptual processes of proximity ratings derived from similarity judgements is questionable and by no means demonstrated. On the other hand, the dependence of proximity ratings derived from confusion matrices upon the particular kind of stress used to generate analysable error patterns is well established.

Current research has not developed to the point where perceptual distances are predictable as performance characteristics of some model that samples as input certain critical acoustic parameters of the signal. This is arguably the goal towards which future research ought to be directed.

CHAPTER IV

ELEMENTS OF MULTIDIMENSIONAL SCALING

In its raw form the proximity matrix generally provides few clues about the the underlying causative factors responsible for the very apparent variation in the values of the off-diagonal elements. As Shepard (1963), one of the pioneering contributors to methodology in this field, has pointed out:

> Man's information processing system...is notoriously unable to discern any pattern in an array of numbers by inspection alone. Therefore ... we must first supplement this natural processing system, with artificial machinery more specifically designed for the task at hand; namely the extraction of implicit structure underlying the explicit but bewilderingly enormous array of numbers [p.33]

It has only been with the comparatively recent development of modern mathematical methods of data reduction such as hierachical clustering (Johnson, 1967), factor analysis (Harman, 1967), and multidimensional scaling (Torgerson, 1958; Shepard, 1962; Kruskal, 1964a; 1964b) that it has become possible to gain access to the presumed latent structure in the raw proximity matrix.

The family of procedures known as multidimensional scaling (MDS) - and the same may be said of factor analysis - attempts to achieve a conceptually useful representation of the data matrix by interpreting the perceptual proximity scores as distances (or more precisly, as some function of distances) in a multidimensional space. The space is generally, though not necessarily, taken to be Euclidean. The actual dimensionality of the space is assumed to be unknown and, together with the distances, it is one of the parameters to be determined by the computational algorithm. The dimensionality of the best fitting representation for the set of objects (phonemic targets in this case, treated as points in a perceptual space) is important for the interpretation of the derived perceptual configuration. The minimum number of orthogonal dimensions necessary to adequately represent the optimal configuration of interpoint distances is indicative of the minimum number of independent variables required to account for the variation of the scores in the proximity matrix.

MDS as a representational scheme favours (but does not dictate) a perceptual model in which the phonemic targets are recognized on the basis of a small number of independent (or near independent) scalar features, which may or may not have readily discernible correlates in the physical signal. If the proximity matrix is mapped into a configuration of points in a Euclidean space, the interpoint distances are invariant under rigid rotation, uniform stretching, and placement of the origin of the reference axes. Given such a mapping, the commonly employed research strategy is to seek a unique rotation of the reference axes which is (a) readily interpretable and (b) supported by independent evidence.

Axis rotation and factor interpretation are central preoccupations in MDS and factor analytic research, but at this point more needs to be said about the model and computational algorithms employed in MDS.

The computational algorithms used in the more powerful, so-called "non metric" MDS procedures (Kruskal, 1964; Torgerson and Young, 1967) strive to attain a configuration of n points that optimizes a monotonic best fit of the derived interpoint distances to the pairwise proximity ratings in a space of minimum dimensionality. Characteristically, the computation begins by setting up an arbitrary configuration of points of some specified dimensionality (r, where $r<n$). The pairwise proximities read from the input proximity matrix are then rank-ordered from smallest to largest. This ranking serves as the criterion to which the algorithm strives to monotonically match the initially arbitrary set of interpoint distances. On each iteration of the algorithm the points of the configuration are shifted around by a small amount in the general direction of a more adequate solution. A measure of the degree of monotonic best fit between the derived distances and the rank ordered proximities is computed after each successive adjustment of the distances. Various measures of best fit have been proposed but they all tend to behave in a very similar fashion (Young, 1970). Kruskal's measure of "stress" is the most commonly used estimate of goodness or badness of fit and also the best researched. It is quite

[illegible] the least squares measure of fit used in re[illegible] analysis:

$$\text{Stress} = \sqrt{\frac{(d_{jk} - \hat{d}_{jk})^2}{\Sigma d_{jk}}}$$

where d_{jk}= derived distance between points j and k

$\hat{d}_{jk}$= monotonic best fitting distance with respect to the proximity p_{jk} .

The iterative convergence on the best fitting configuration ceases when no further significant improvements in stress can be achieved.

Stress values vary between 0 (perfect fit) and 1 (no fit at all), or may be expressed as percentages. Kruskal (1964) provides a set of descriptive labels which he suggests the user employ as a rough guideline for evaluating the adequacy of the obtained monotonic matching. Systematic research with artificial data (Klhar, 1969; Young, 1970; Sherman, 1972) suggests that Kruskal's criterion is too conservative, and that the evaluation of stress is a complicated matter requiring that account be taken of the number of objects scaled, the dimensionality of the final solution, and the anticipated error of estimation of the proximity scores.

The criterion of monotonic matching between the proximities and the derived distances is an appealing one. There are usually no grounds for anticipating, beyond rank-ordering, what the precise form of the relationship between the proximity scores and the hypothetical distances will or should be. The actual form of the relationship is obtained as a by-product of the computation from the final, monotonic best fitting curve of the derived distances plotted against the proximities.

The earlier, so called "metric" variety of MDS (see Torgerson, 1958) required the stronger assumption of linearity between the proximities and the final obtained distances. Where it is apparent from a non-metric solution that the assumption of linearity is not too badly violated, or where the proximities can easily be transformed to render the relationship linear, it is advisable to employ the metric MDS technique. Its computational algorithm is quite different from the non-metric routines. If essentially the same configuration is recovered with the metric technique the investigator can be more confident in the uniqueness and stability of his final solution.

Estimation of Dimensionality

MDS algorithms usually yield solutions in a range of dimensions specified by the user. Mathematically, the solution of lowest dimensionality that adequately represents the proximity matrix is preferable because it is the most strongly determined. An argument for minimum dimensionality can also be made in terms of descriptive economy which boils down to: Why postulate the operation of a greater number of variables than you need? However the fact that stress values usually decrease as the dimensionality of the representational space increases makes the choice of the optimal dimensionality in many cases rather difficult. Customary practice is to plot the function of stress against dimensionality and look for a point of diminishing returns beyond which increasing the dimensionality does not significantly improve stress ratings, in that each increase accounts for the same improvement, and may be considered to be fitting only the noise in the data. But perhaps the major criterion employed for deciding on the dimensionality rests (in practice at least) on the interpretabliity of the rotated reference axes. The investigator will be reluctant to extract from his data more factors than those for which he can find a plausible interpretation.

Measurement Error and Configuration Recoverability

A fundamental question for the researcher is how confident can he be that the the apparent perceptual structure reflected in an obtained MDS solution reflects real, latent structure inherent in the proximity matrix? Put in the form of the null hypothesis: How can he be sure that the variation in the off-diagonal values of the proximity matrix is not randomly generated? After all, any matrix of substantially non-zero entries will yield some kind of MDS solution. The simplest and safest answer to this question lies in the replication of findings over independent sets of data. Another guide is the stress of the final solution. Klhar (1969), and Stenson and Knoll (1969) obtained distributions of stress values for completely random matrices representing various numbers of perceptual objects scaled over a range of dimensions. They found that where the dimensionality of a solution is small with respect to _n_, the number of objects scaled, and _n_>10, the standard deviation of the corresponding distribution of stress values is very small. The investigator can use these results to construct rough confidence limits for testing the null hypothesis that _no_ latent structure exists in his proximity matrix.

In practice it is unlikely that no structure exists in the data matrix, but rather, such structure as does exist will be overlaid with a certain amount of measurement error. The amount of error may be difficult to estimate. However, it is useful for the user of MDS to know under what conditions the solutions yielded by an MDS routine are robust under the assumption of measurement error. The standard procedure for testing a MDS algorithm is to construct a particular configuration of _n_ points in _r_ dimensions, subject the interpoint distances to some arbitrary monotonic transformation, treat these transformed distances as proximities, and see how well the original structure can be recovered by the algorithm. Young (1969) investigated the effect upon configuration recovery of adding different levels of random error to the interpoint distances, prior to the monotonic transformation to "proximity" scores. Even at the highest level of error studied (where the variance of the random error component = 35% of variance of interpoint distances) recovery of the original configuration was good (correlation between original and derived interpoint distances >.8), providing the ratio between the number of objects scaled and the dimensionality of the configuration was fairly high (>3). Where the dimensionality of the representational space is low compared to the number of objects being scaled, the solution is strongly "overdetermined" (Shepard, 1962). This is the reason why only the weak assumption of monotonic relationship is necessary between the proximity scores and the derived distances and why the algorithm seems to be quite robust against the effects of error variation in the proximity scores.

The Problem of Rotation and the INDSCAL model

Standard "metric" and "non-metric" MDS and factor analysis routines employ a Euclidean spatial metric and yield solutions which are unique up to the point of axis rotation. This may be regarded as an asset or a liability depending on one's theoretical predisposition. The rotational indeterminancy of standard Euclidean MDS and factor analysis provides the investigator with greater freedom in arriving at a theoretically satisfying solution. However, when the obtained dimensionality of the solution is comparatively high, selecting "the best" rotation for the reference axes can be very diffuclt. Two or more substantially different interpretations of the scaling configuration may be suggested (each of which constitute mathematically equivalent representations of the data matrix).

In the area of factor analysis, general criteria (such as Thurstone's (1947) "simple structure") have been proposed and corresponding computational routines developed that will take an initially "arbitrary" set of co-ordinate axes and transform them into a new set of reference axes that optimally meet the rotational criterion. The "simple structure" criterion has proved popular because, by simplifying the pattern of factor loadings on the input variables, it has tended to yield solutions that are readily interpretable in terms of those variables. Rotation in accordance with the principal components criterion, which locates the reference axes in accordance with eigenvectors that encompass successively diminishing amounts of the variance in the configuration, generally leads to less readily interpretable factors than the simple structure criterion. The historical debate between the "American" and the "British" schools of factor analysis over the factorial structure of human ability involves just this point.

In recent years, the development of three-mode methods of MDS have provided the possibility of obtaining rotationally unique Euclidean scaling solutions (Carroll & Chang's INDSCAL, 1970; Harshman's PARAFAC, 1970). Whether or not these procedures yield "explanatory" as distinct from merely "descriptive" factors (see Harshman, 1970) is a moot point. The model is able to attain rotational uniqueness by making certain (quite strong) assumptions about the nature of the variability across the third mode, which is usually taken to be individuals.

With the INDSCAL method, variation in the third mode is restricted to the application of a weighting factor (w) applied to a given individual i on a given dimension t. Conceptually this weighting factor may be interpreted as the relative salience a particular underlying perceptual dimension has for a given individual. The model for the interpoint distance between two stimuli j and k for individual i is:

$$d^{i}_{jk} = [\sum^{t} w_{it} (x^{i}_{jt} - x^{i}_{kt})^{2}]^{\frac{1}{2}}$$

> A principal disadvantage of the method is that it is limited to the case in which individual subject spaces are related by linear transformations of a common space. Even the linear transformations allowed are not general, but are restricted to those given by diagonal transformation matrices. The method may require too many dimensions in cases where the perceptual spaces represent nonlinear distortions of a common space (or where more general linear transformations are required) [Carroll & Chang, 1970, p.316].

In practice, this assumption may not prove to be too restrictive. Carroll and Chang (1973) provide, among other illustrations, one with a set of data on similarity ratings of synthetic auditory stimuli (Bricker, Pruzansky, & Mc Dermott, 1968) which is relevant to the subject of this paper. The INDSCAL analysis of subjects' similarity ratings yielded a perceptual configuration in essential agreement with that which would have been predicted from the physical parametric variations used in constructing the set of stimuli. This alone is no more than might be expected from a good standard MDS routine. What was remarkable however, was that the placement of the reference axes corresponded (without rotation) in a one-to-one fashion with the *a priori* physical dimensions manipulated in the synthesis of the stimuli.

Carroll and Chang argue that the proof of the pudding is in the eating:

> In cases where set of *a priori* physical or theoretical dimensions were known, the recovered (unrotated) dimensions have always (to date) corresponded to them in essentially one to one fashion. We therefore argue that it is appropriate to analyse data in terms of this very strong and specific model, and that only if this model fails to fit the data adequately should one have recourse to a more general model [p.285].

On the other hand, in an exploratory investigation it is undesirable to have the interpretability of the solution too dependent on strong assumptions of the scaling model. In this context the rotationally non-unique, standard MDS procedures have much to recommend them, at least in the early stages of factor identification.

The Spatial Metric

Kruskal (1964) introduced the option of scaling in non-Euclidean spaces. His MDS routine is applicable to the general class of Minkowski spatial metrics of which the Euclidean and the "city block" metrics are special cases. The formula for calculating interpoint distances for the general Minkowski spatial metric is:

$$d_{jk} = [\sum^{t} (x_{jt} - x_{kt})^{n}]^{\frac{1}{n}}$$

where n = some positive real number.

In the case of the "city block" metric where the exponent (n) is 1.0, the distance between two points is simply the sum of the absolute differences of the point projections on the reference axes. In other words, equal weight is given to the differences on each dimension, regardless of their relative magnitude, in the determination of the interpoint distances. It is intuitively evident that, as n increases, progressively more weight is given to the larger differences on the projection axes. In the limiting case where $n \rightarrow \infty$, only the largest difference on any dimension contributes to the interpoint distance d_{jk}.

In this family of spatial metrics, only the Euclidean preserves invariance of the interpoint distances over axis rotation. The question naturally arises as to what kind of spatial metric is appropriate for the representation of perceptual distances? This is very difficult to answer. One approach to this problem has utilized arrays of simple physical stimuli where the dimensions and the levels of variation within a dimension are fairly clearly established. Interpoint distances are derived, usually on the basis of a Euclidean or a city block metric from a set of perceptual proximity measures of one kind or another. The test of the adequacy of a given spatial metric is its ability to yield perceptual distances that closely conform with the physical parameters of the stimuli.

Several studies along these lines (Attneave, 1950; Torgerson, 1952, 1965; Shepard, 1964; Hyman & Well, 1967, 1968; Garner and Felfoldy, 1970) have resulted in a categorical distinction between "analysable" and "non-analysable" stimuli. In the former case, the underlying perceptual dimensions are obvious and distinct to the subject, such as when the stimulus material consists of simple geometric patterns varying in, for example, size and angle of inclination. In the latter case, the underlying perceptual dimensions are not distinct and obvious to the subject but qualitatively "integrated". More elementary perceptual objects, such as colour sensation, or simple auditory or tactile displays, belong to this class of stimuli. Only the "unanalysable" stimulus displays have been found, by the above methodology, to scale adequately in accordance with the Euclidean metric. "Analysable" stimuli have been found to conform better to the city block space, or, because of some gross violation of the triangular inequality, to not map satisfactorily into any well understood geometric representation (Shepard, 1964).

On the basis of these studies it may be concluded that scaling in accordance with a Euclidean metric is appropriate for elementary perceptual targets such as phonemes embedded in a constant syllabic frame, where the underlying perceptual dimensions are non-obvious. However, a cautionary note is warranted, because it seems precisely in those cases where the underlying perceptual dimensions are "unanalysable" that the method of vindicating a spatial metric by showing that it yields a perceptual configuration in close agreement with the supposed relevant physical parameters of the stimuli, is most questionable.

Another approach to this problem that has been tried is that of Terbeek and Harshman (1971) who were led to question the validity of the Euclidean metric for vowel perception. They consistently found an extra and interpretively intransigent dimension in their scaling solutions, loadings on which turned out to be highly predictable by a simple (non-linear) function of two other dimensions in the scaling solution. These results were consistent with a hypothesis of spatial curvature which could create conditions leading to the extraction of an extra, spurious dimension when the perceptual distances were inappropriately represented in a Euclidean spatial metric.

No strong theoretical considerations support the choice of a Euclidean over some alternative spatial metric, but for pragmatic reasons the Euclidean metric has much to recommend it. The properties of a Euclidean model are well understood. Only the more recent "non-metric" varieties of MDS allow for other than Euclidean solutions and therefore cross-technique comparisons to test the stability of a derived configuration can only be made within the Euclidean framework. Also, parameter testing experiments with synthetic data (Sherman, 1972) have suggested that the choice of the spatial metric significantly affects the retrievabliity of a configuration only when the dimensionality of the solution is correctly identified.

CHAPTER V

EXPERIMENTAL RESULTS

In this chapter a sequence of four inter-related experiments is described, in which the general aim was to isolate and describe the most salient perceptual dimensions involved in the recognition of a selected set of English consonantal phonemes in open syllable position. In each case, proximity matrices were obtained to characterise the perceptual relationships between all items in the stimulus set. These were derived from either direct or indirect similarity judgements. A variety of multivariate scaling procedures - "non-metric" MDS (Kruskal, 1964); "metric" MDS (Torgerson, 1958); Principle Components Factor Analysis (Harman, 1967); and hierachical clustering (Veldman, 1967) - were employed, chiefly to test the robustness of derived solutions under conceptually related but computationally diverse analytical techniques. In all cases, group rather than individual proximity matrices were analysed. [illegible]is was necessitated by practical considerations of subject availability which in turn influenced the choice of data collection procedures. However, it seemed to be a reasonably safe assumption that on a basic perceptual task of the kind involved in these experiments, intersubject variabliity should not be of significant theoretical interest. But this remains an untested assumption that ought to be considered in future investigations.

Experiment I was an exploratory investigation employing MDS of direct similarity judgements of English consonants embedded in a /Ca/ syllabic frame. Two factors, thought to be basic auditory features that could be used to differentiate the stimuli, were tentatively identified. Experiment II was a larger study concerned with replicating the findings of Experiment I and determining the stability of the obtained perceptual configuration when a phonetically quite different vowel is employed in the constant (carrier) /CV/ frame. Results indicated that while the basic configuration is maintained, some significant (systematic) perturbation took place as a function of the phonetic quality of the vowel. Experiment III constituted an attempt to determine whether, on the basis of the factors discerned in Experiments I and II, it would be possible to predict the locations in perceptual space, of consonants not included in the original scaling set. In Experiment IV a proximity matrix for the stimuli used in Experiments I and II was generated from indirect estimates of perceptual similarity based on semantic scaling. The rationale for this experiment was simply that if the same perceptual dimensions successfully describe proximity matrices generated by two quite different, but theoretically well motivated data bases, then the case for the psychological reality of such perceptual dimensions is considerably strengthened.

Subjects

The subjects for these experiments were drawn from several sources:

(i) Students enrolled in general introductory linguistics courses at the University of Alberta; (Experiment I, n=27; Experiment IV, n=30).

(ii) First and second year junior college students from the Red Deer and Medicine Hat Colleges in Alberta (Experiment II, n=90).

(iii) Students enrolled in an evening course in English Literature at the University of Alberta (Experiment III, n=22). The great majority of the subjects may be described as naive with respect to formal phonetic training. All subjects included in the analysis were native speakers of English, where "native speaker" is defined as having "used English as your major language since you were five years of age."

Experiment I

In this investigation 12 English consonants /p,b,t,d,č,s,š,z,n,m,l,h/ embedded in a /Ca/ syllabic frame were scaled for perceptual similarity employing a modified method of triadic comparisons. The obtained group proximity matrix was subjected to Kruskal's method of MDS and also to a hierachical clustering analysis.

Choice of Stimuli

Because the total number of trials in an experiment employing triadic comparisons increases very quickly as a function of the number of objects scaled, this study was constrained to operate with a subset of the consonantal inventory. An attempt was made, admittedly on intuitive grounds, to make the chosen set representative of the range of perceived auditory variability encountered in the complete set of consonantal phonemic targets.

Method

The modified method of triadic comparisons used in this experiment is an adaptation of the method of "complete triadic comparisons" outlined in Torgerson (1958). Consider the set of all possible three-way (triadic) combinations of n perceptual objects to be scaled. Subjects are presented with one triad at a time, drawn randomly from the $(n!/(n-3)!3!)$ triadic combinations. On each trial they are instructed to assign a rating of 2 to the pair of stimuli out of the 3 which are most alike, and a rating of 0 to the two which are least alike. The remaining unchosen pair is assumed to take a proximity rating somewhere between the "most alike" and the "least alike" pairs and is automatically scored 1. Hence over all triadic combinations, pairs most frequently chosen as "most alike" will emerge with a high score and those mainly chosen "least alike" with a low one on a scale of overall similarity. This method, which requires that the objects be presented three at a time, works better for visual than auditory stimuli since the latter cannot be scanned, but must be distributed in time. To mitigate possible problems of trace decay in short-term auditory memory, each triad was decomposed into three simple pairwise judgements. To illustrate the strategy, consider the possible triad:

ča

pa da

from which the following pairs may be formed:

A	B
pa - ča	pa - da
ča - pa	ča - da
da - pa	da - ča

Each element of the triad acts once in the three pairwise presentations as the standard for a simple perceptual judgement: Which pair of syllables (A or B) sound most alike?

The total set of pairwise presentations generated from all triadic combinations were recorded by the experimenter in a randomized order, with appropriate pause intervals placed between the stimuli. A single speaker (the experimenter) was used for recording all the stimuli. Because of the large number of trials in the overall experiment (1,980) each subject judged only a portion of the total set of comparisons. Twenty-seven subjects were used, divided into nine groups assigned to systematically overlapping blocks of trials (220 per block). Each testing session lasted approximately 30 minutes. Subjects checked their responses (either the first or second pair for any trial) on an optically scorable IBM answer sheet.

Procedure

A group testing situation was used. After an informal introduction to the purpose of the experiment and a general description of the experimental task, subjects were orally presented with the following instructions:

> You will hear pairs of syllables presented over the loudspeaker. Pairs such as: (pause) pa - ca (pause) pa - ba (pause). Your task is to ask yourself which of the two pairs sounds more alike: the first or the second pair? In this case most people would probably say that the pair pa - ba sounds more alike than the pair pa - ca. Pa - ba was the second of the two pairs, so in this case you would mark the second alternative on your answer sheet. On the other hand, if the syllables in the first pair sound more alike than the syllables in the second pair, you would put a mark in column one of your answer sheet. On some items you will have difficulty deciding which of the two pairs sounds most alike. However, we would like you to choose one of them even if it seems sometimes like "guesswork." Don't spend too long making up your mind. It's your first impression that we are interested in. In deciding which of the two pairs is most alike just go on the sound of the syllables, not on how they might have been produced... (further instructions about recording responses

on answer sheets)... Any questions?

Scoring the Responses

A group proximity matrix was accumulated in the following manner: An initially empty 12 x 12 scoring array was set up in which the rows represented the 12 syllables (or phonemes) as standards and the 12 columns represented possible responses. On a given trial , one of the two possible response syllables (the columns of the scoring matrix) may be associated with a particular syllable (or phoneme) as standard (the row elements). A "1" was added to the appropriate row and column intersection of the scoring matrix each time a particular standard and response were associated by being chosen as the "more similar" pair. When accumulated over all the experimental trials, the scoring matrix indicates the relative frequency with which particular syllable pairs are chosen as "more similar" than all other pairs. The accumulated scoring matrix is approximately (but not precisely) symmetrical. To meet the necessary assumption of symmetry and possibly improve the stability of the similarity scores, the corresponding off-diagonal elements of the cumulative scoring matrix were summed to produce a symmetrical proximity matrix (Table 5.1).

TABLE 5.1

PROXIMITY MATRIX EXPERIMENT I

	p	b	t	d	č	s	š	h	z	m	n
b	101										
t	77	66									
d	76	90	107								
č	32	46	77	47							
s	40	40	54	44	72						
š	45	28	51	38	104	82					
h	63	75	56	57	64	53	63				
z	35	48	43	53	64	101	82	45			
m	75	73	29	68	33	42	36	72	39		
n	60	63	50	76	33	41	39	78	46	105	
l	61	68	47	63	29	49	45	78	50	90	92

Results

The next step was to determine the optimal spatial configuration for the proximity matrix. For reasons given earlier, the Euclidean metric was chosen as a basis for computing the interpoint distances in spaces ranging from one to nine dimensions.

Figure 5.1 shows the adequacy of monotonic matching (stress) as a function of the dimensionality of the solution. Several computational runs from different starting dimensions were made to help ensure that the preferred dimensionality would emerge clearly. According to the

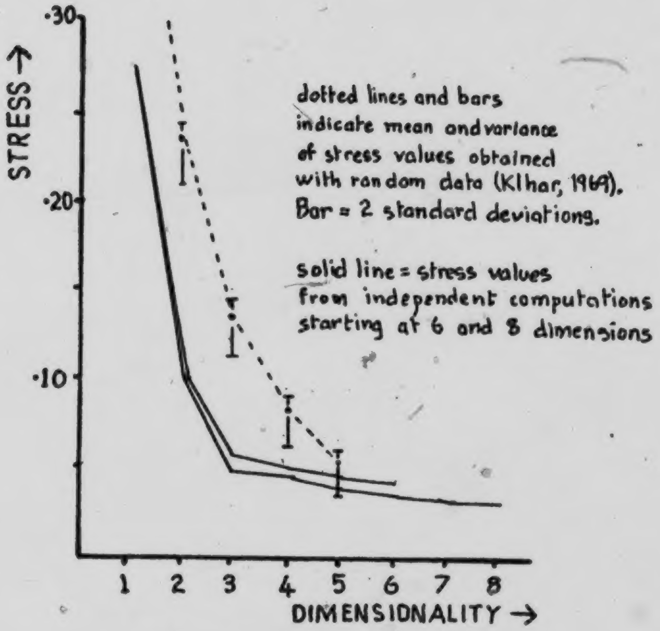


Fig. 5.1 Stress X Dimensionality
Experiment I

criterion of minimal stress with minimal dimensionality, the graph indicates that a three-dimensional solution (with a "good" stress of 5% by Kruskal's conservative rating) is the preferred solution, although the two dimensional configuration (stress=10%) should be considered in attempting to arrive at an adequately interpretable solution. Figure 5.1 also shows confidence limits, based on Klhar (1969), for the null hypothesis of no latent structure in a proximity matrix for 12 objects scaled in 1 to 5 dimensions. Clearly the null hypothesis may be rejected. However, with only 12 objects scaled, one would be disinclined to accept a dimensionality higher than four.

The three-dimensional configuration was examined first (see Figure 5.2). The orientation of the reference axes is

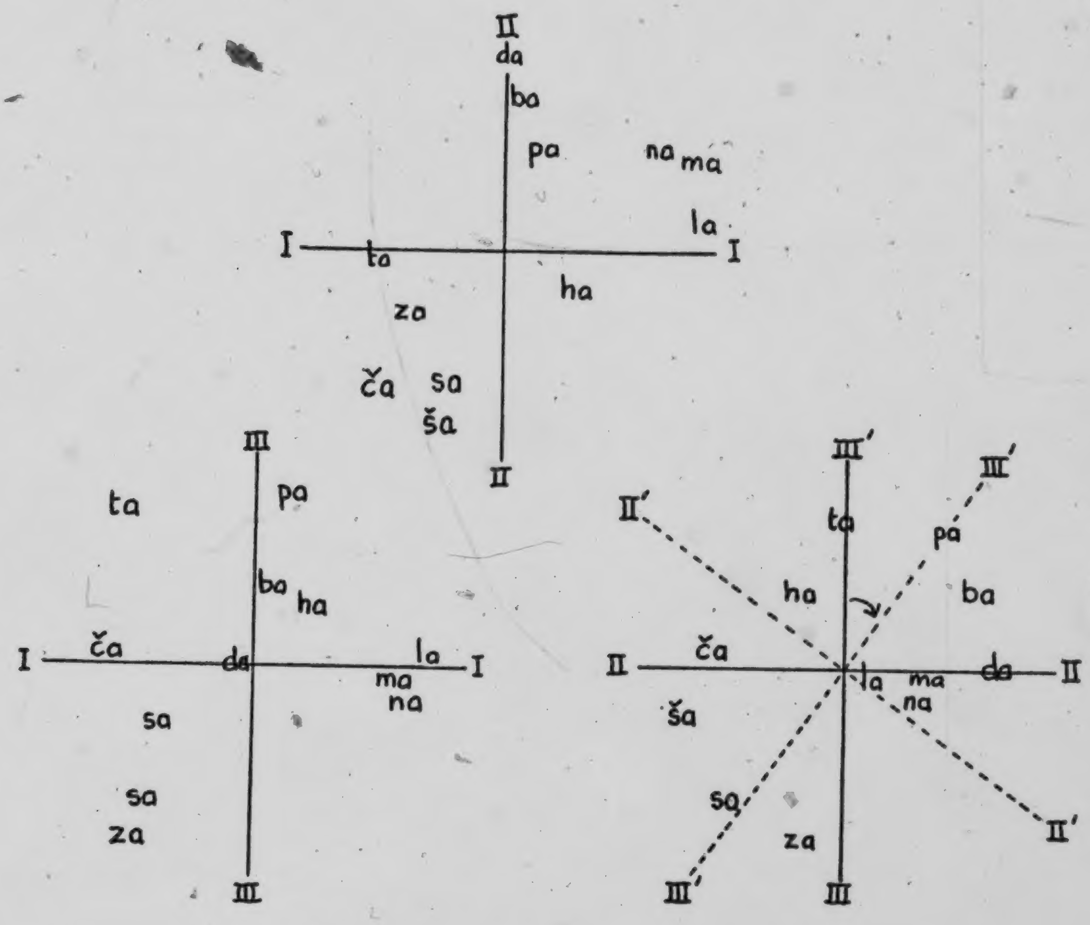


Fig. 5.2 Three Dimensional Configuration Experiment I Kruskal Scaling

arbitrary, but two of the three orthogonal axes in the three-dimensional solution seemed to be readily interpretable.

Dimension I locates the resonants /la,ma,na/ on one pole and the affricate /ča/, the voicless stop /ta/ and the fricatives /za,sa,ša/ on the other. It opposes sounds with harmonic and formant structure and a "musical" quality to those with spectral energy distributions characteristic of a turbulent noise source. For want of better teminology this was tentatively labelled the "resonance - hiss" dimension, or simply "resonance."

Unrotated Dimension III /ta,pa,ba/ versus /za,sa/ could be construed as a temporal factor: the duration of the consonantal portion of the syllable. Consistent with this interpretation, the nasals, the lateral, and the affricate are located in the middle of this continuum. However, inconsistent with this interpretation, the stop /da/ instead of loading positively with the other stops on unrotated Dimension III takes an intermediary value together with /ča,ma,la,na/.

Unrotated Dimension II is even more difficult to conceptualize. What plausible auditory, acoustic, or articulatory scale would oppose the voiced stops /da,ba/ to the sibilants /ča,sa/ with medial values assigned to /ta,za,la/?

A considerable improvement in interpretability of the three-dimensional solution may be obtained with a 45 degree clockwise rotation of the reference axes in the DII-DIII plane (see Figure 5.2). The notable discrepency that clouded the interpretation of Dimension III as a temporal factor is removed. Dimension III' clearly opposes the stops to the continuants. Dimension II' may be interpreted as a voicing dimension. All the voiceless consonants have positive loadings on this (rotated) axis and all the voiced consonants are negatively loaded.

Examination of the two-dimensional solution revealed an interesting parallel with the three-dimensional configuration. The two dimensions of the three-dimensional solution - tentatively identified as "duration" and "resonance" - are clearly suggested by the two-dimensional configuration (see Figure 5.3). The two-dimensional configuration is also useful for representing the three major perceptual groupings that emerged when the proximity matrix was subjected to a hierachical clustering algorithm (Veldman, 1967). It is interesting that these three statistically derived clusters correspond to the traditional manner of articulation categories which are represented in the scaling set of phonemes: the stops, the resonants, and the sibilants.

Even within the major clusters, the target phonemes appear to be differentiated in the appropriate manner by the two inferred dimensions of sonority and duration, which would suggest these are scalar rather than categorical perceptual features. However, before these results could reasonably sustain the weight of such an interpretation a number of questions about their reliability, replicability, and range of applicability needs to be answered.

Experiment II

Experiment II constituted a replication and extension

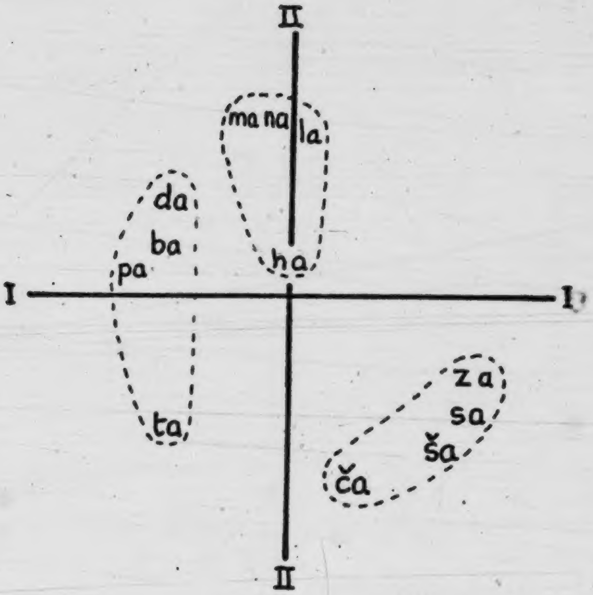


Fig. 5.3 Two Dimensional Solution
Experiment II Kruskal Scaling

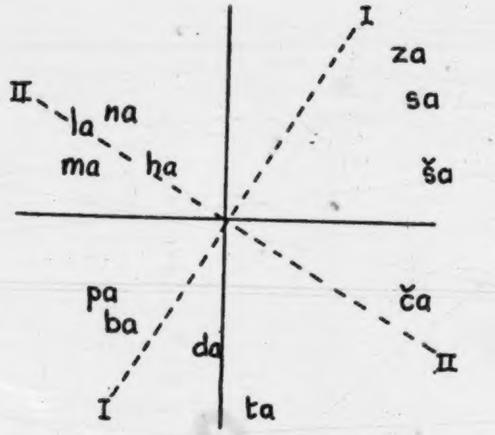
of the first experiment, with MDS of triadic comparisons of perceptual similarity. Two sets of stimuli were scaled in interpolated trials. One set was the same as that employed in Experiment I - 12 consonantal phones embedded in a /Ca/ syllabic frame. The second set comprised the same 12 consonants but combined with the vowel /i/. The 1,320 trials resulting from a randomized combination of the two sets of 660 (X - A X - B) pairwise comparisons were divided into 6 experimental blocks of 220 trials. There were 15 to 24 subjects per block of trials. The experimental procedure was identical with that of Experiment I, except for the method of constructing the stimulus tapes. Instead of separately recording each trial, the basic set of 24 CV syllables were recorded once, digitalized by a program written for the PDP

12 computer (Roszypal, 1973) at a sampling rate of 15kHz and stored on LINC tape. A second program constructed the experimental tape from the LINK-stored stimuli.

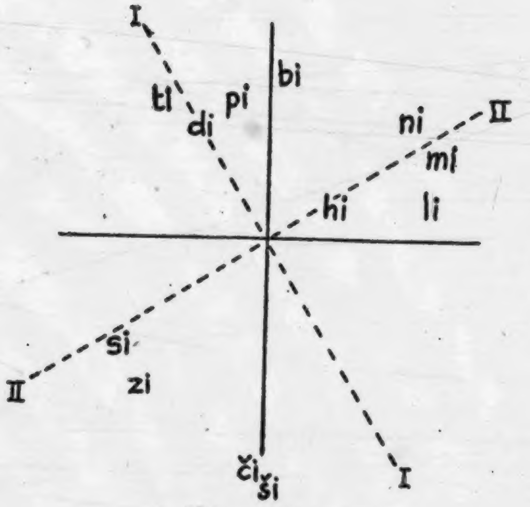
Results

Group proximity matrices for both the /Ca/ and the /Ci/ sets were accumulated (Appendix D) and the data were scaled in from one to four dimensions by Kruskal's method. The plot of stress against dimensionality (Appendix D) did not reveal a clearly preferred dimensionality for the /Ca/ set. For the /Ci/ set the scree test suggested that the two dimensional solution (stress = 9%) was preferable.

The two and three-dimensional Kruskal solutions for the /Ca/ set were compared with the two and three-dimensional solutions of Experiment I to determine the degree of replicability of the scaling configurations over independent data sets and to see whether the same hypothetical dimensions of "resonance-hiss", "duration", and "voicing" could be maintained. For the two-dimensional solution, a comparison of Figures 5.4a and 5.3 shows a satisfactory replication of the results of Experiment I. When the axes of Figure 5.4a are rotated (graphically) to conform to the orientation of Figure 5.3 it will be noted that the loadings of the sounds on each of the dimensions are highly similar. There is some discrepency with respect to the stop consonants on the "hiss - resonance" dimension. It can be



(a) /Ca/ set



(b) /Ci/ set

Fig. 5.4 Two Dimensional Solution
Experiment II Kruskal Scaling

seen that /h/ shows a slightly stronger tendency to cluster with the resonants. Rather than attempt to interpret these minor discrepencies, which could have their origin in the different stimulus sets used in the two experiments, it seems better to regard them as indicative of limits on the accurracy of resolution obtainable with the data collection procedures used in these experiments.

The two most prominent dimensions found in the first experiment, "resonance-hiss" and "duration", are clearly apparent in the two dimensional solution of Experiment II. They may also be discerned in the three-dimensional solution for the /Ca/ set (Appendix D). However, the "voicing" dimension, which was clearly the weakest dimension in the results of Experiment I, failed to emerge in the three-dimensional solution for the /Ca/ set in Experiment II. (Though there is an equivocal suggestion of a voicing dimension: see Appendix D.) The reason for this failure to yield a clearly discernable voicing dimension in Experiment II may lie in differences between the subject pools of the two experiments. Most of the subjects employed in Experiment I had some knowledge of linguistics and therefore probably some familiarity with phonetic description. As the voicing feature is particularly prominent in traditional phonetic classifications and pedagogic illustrations of phonological rules, it is plausible to suggest that its clear presence in the data of Experiment I but apparent absence in Experiment II simply reflected differential linguistic training of the two groups.

The two replicable dimensions of "resonance" and "duration" found with the /Ca/ stimulus sets are apparent also in the scaling configurations for the /Ci/ set of

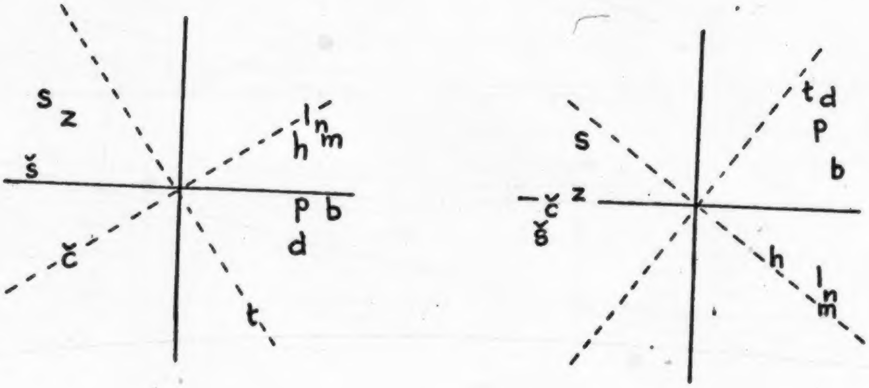
Experiment II. They emerge most clearly in the two-dimensional solution (see Figure 5.4b), but are also discernable in the three-dimensional solution (Appendix D) where, again, there is a weak suggestion of a "voicing" dimension.

In both the two and the three-dimensional solutions for the /Ci/ set, the clarity of the hypothesised temporal factor is obscured by a transposition of the rank ordering of the sibilants on this dimension. To further clarify the question of whether this shift is attributable to structural differences in the two proximity matrices and not to a computational indeterminancy associated with the scaling algorithm, the data of Experiment II were subjected to Torgerson (1958) scaling and a Principal Components factor analysis program, with options of varimax and oblique axis rotation (Program DERS:FACT04 in the Division of Educational Research program library, University of Alberta).

The input to the Torgerson scaling program (DERS:SCAL05) was a comparative distance matrix based upon z score transformations of "proportion of choice" scores (see Torgerson, 1958). The program estimates absolute distances for spaces of varying dimensionality by an iterative procedure based on Messick and Abelson (1956). For major steps in the computation of the interpoint distances and details of the scaling solution see Appendix E.

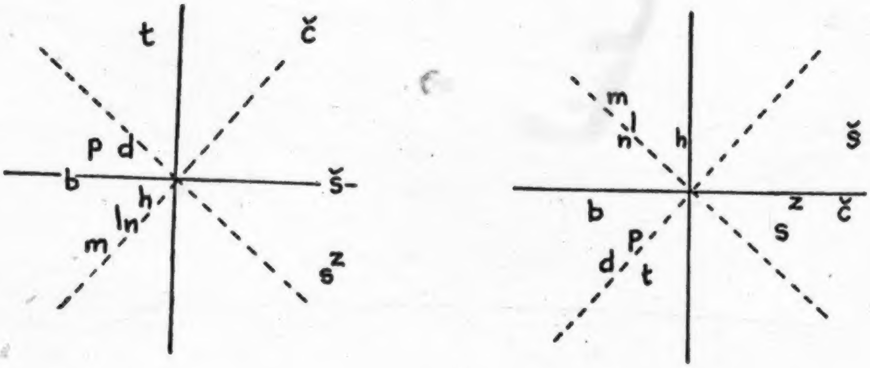
For the factor analysis, the rows of the proximity matrices for the Kruskal analysis (Appendix B) were intercorrelated. Each of the two resulting correlation matrices (Appendix F) represents a set of indices of similarity: namely, the similarity of any given syllable as a standard (i.e., "X" in the X - A X - B trial) with any other syllable as a standard in the set of 12 stimuli. Details of the factor analysis are given in Appendix F).

Both the Torgerson and the Factor Analysis programs yield, in the first instance a "principle axis" type of solution, i.e., a configuration in which the first axis is located so that its object loadings account for the maximum amount of the common variance in the data and subsequent axes account for progressively diminishing amounts of variance. No uniquely prefered dimensionality emerged from the four analyses. There is, however, considerable consistency between the results of the two scaling methods. Figure 5.5 shows a plot of the first two principle components of the four-dimensional solutions for the /Ca/ and the /Ci/ data sets as yielded by both the Torgerson and Factor Analysis programs. the four-dimensional solution is chosen as being the highest one might reasonably anticipate for the relatively small number of objects scaled. It is notable that for all four solutions the first principal component which emerges clearly separates the fricatives and the affricate from the stops and the resonants, accounting for approximately 50% of the total variance. By rotating the reference axes approximately 30 degrees from the first two



Factor Analysis
/Ca/ set

Factor Analysis
/Ci/ set



Torgerson Scaling
/Ca/ set

Torgerson Scaling
/Ci/ set

Fig. 5.5 First Two Principle Factors
Experiment II Torgerson Scaling and Factor Analysis

principal axes one can obtain a two factor solution (accounting for about 75% of the common variance) in substantial agreement with the Kruskal solution (Figure 5.4a). Moreover, the transposition of the sibilants on the "temporal" axis for the /Ci/ set noted in the Kruskal solution is apparent also in the Torgerson scaling and the Factor Analysis. Possible reasons for this perturbation in the perceptual configuration are discussed in the following chapter.

The varimax and the oblique rotations of the factor analysis are of considerable interest. For the /Ca/ set, with both the varimax and the oblique solutions, the first two axes suggest the "duration" and "resonance - hiss" factors respectively, (i.e., Factor I opposes the stops /t,d,b,p/ (in that order) to the fricatives /s,z,š/; Factor II has positive loadings on the resonants /n,m,l/ and significant negative loadings on /s,š,č,z/).

The first two varimax and oblimax factors for the /Ci/ set also yield the dimensions "duration" and "resonance" respectively - providing one can accept that a shift in the relative duration amongst the sibilants caused the perturbation in the perceptual configuration which was noted earlier. The pattern of loadings on Factors III and IV are consistent across both orthogonal and oblique rotations but are not readily interpretable. Nor do they agree with Factors III and IV for the /Ca/ set.

Experiment III

If the basis of the subjects' similarity judgements obtained in the two previous experiments is adequately captured by the proposed two factor model, then from a knowledge of their phonetic properties, it should be possible to predict the location in perceptual space of consonants not included in the original scaling set. The new set of stimuli must encompass roughly the same range and type of auditory variability found in the original scaling set if the same perceptual dimensions are to have a chance of emerging from the scaling solution. Also, in order to determine whether factor invariance holds across any two independent scaling studies it is necessary for both sets to share a core of common objects. These requirements are best satisfied by including in the new scaling set a "common core" from the original set which load most highly (and "purely") on the hypothesised underlying factors. Thus /pa/ and /za/ were chosen on the basis of the Kruskal scaling solutions of Experiments I and II, as representing extremes on the "temporal" dimension. Similarly, /ča/ and /la/ were chosen to represent the "resonant - hiss" dimension. With the addition of four "new" sounds /ka,ga,wa,fa/ this comprised the set of stimuli for Experiment III.

With varying degrees of certainty, the "new" sounds should be predictable in the perceptual space of the "old". The /w/ is clearly a strong resonant with a somewhat less abrupt onset and a slightly longer duration than /l/. The stops /k/ and /g/ should clearly cluster with /p/ on the temporal dimension. Their ranking on the resonant dimension is not predictable because Experiments I and II are not particularly consistent on this point. The location of /f/ is somewhat more difficult to predict. Labial frication has a decidedly "softer" quality than alveopalatal frication. It therefore may be expected that /f/ would load closer to the resonant pole of the "resonance - hiss" dimension than /č/ or /z/ (though the voice component in the latter renders this partially debatable). In order to ensure that the labial frication would not be lost in digitalizing the signal, this sound was produced with somewhat heavier emphasis than it might normally obtain in "list reading" the items. For this reason, /f/ - in this experiment at least - should load with the continuants on the temporal dimension.

The eight stimuli were recorded and the experimental items constructed in the same manner as for Experiment II. The scaling procedure was identical with that of the previous experiments.

Results

Again, Kruskal's stress criterion did not clearly favour the two-dimensional solution, but the three-dimensional configuration was difficult to interpret and no higher a dimensionality would be justifiable with such a small number of objects scaled. The two dimensional configuration did however conform well with the predicted scaling solution (see Figure 5.6). The "anchor point" stimuli did not shift relative locations substantially (though /p/ emerges with an uncomfortably high loading on the "resonant" dimension). The locations of the "new" consonants are, more or less, where they "ought" to be.

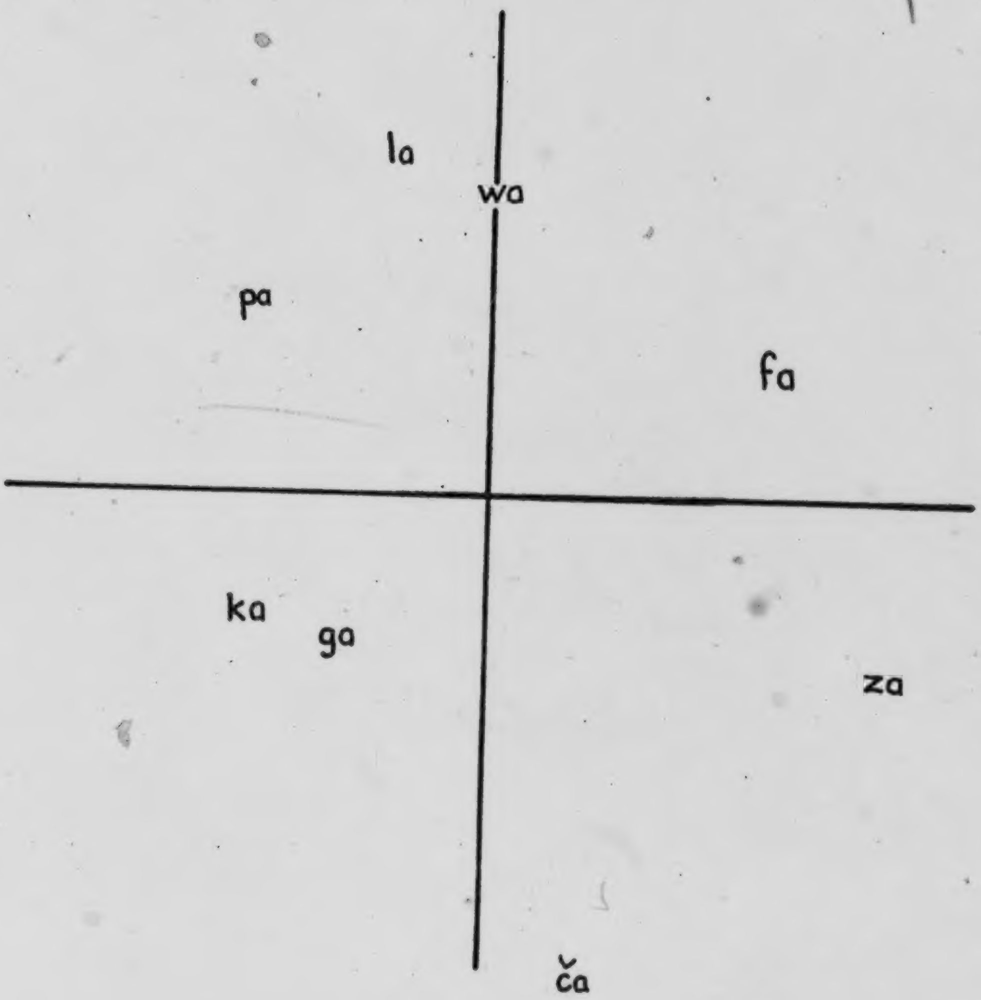


Fig. 5.6 Two Dimensional Kruskal Scaling Solution Experiment III

Experiment IV

Experiment IV was undertaken with the original set of 12 /Ca/ syllables in order to asses the impact of experimental procedures used in constructing the proximity matrix and also to provide additional information for characterising the nature of the factors underlying the derived perceptual configuration. Obviously, if perceptual distances or proximities vary substantially and unpredictably over different but otherwise well motivated techniques of measuring perceptual similarity, then the utility of the whole method is brought into question.

In Experiment IV, an indirect measure of similarity was derived from subjects' ratings of the phonemes (embedded in the same Ca syllabic frame) according to a set of 13 verbal scales thought to be relevant for describing subjective qualities of sounds in general and speech sounds in particular. The scales chosen, mainly formulated on a "most -least" continuum were:

most hissy - least hissy
most vowel like - least vowel like
most bright - least bright
long - short
most clear - least clear
most harsh - least harsh

most distinct - least distinct
high pitch - low pitch
most abrupt - least abrupt
most even - least even
most loud - least loud
most melodious - least melodious
most effortful - least effortful

In selecting these scales, an attempt was made to avoid descriptors that posessed a specialized linguistic or phonetic connotation. (We were not interested in examining the subjects' academic knowledge of phonology or phonetics.) An effort was also made to sample as broadly as possible from the domain of discourse: in this case, "ways of describing sounds." Also, care was taken to represent what were hypothesised to be the underlying factors responsible for the derived perceptual configuration in Experiment I.

Method

Thirty subjects rank-ordered the 12 syllables on each of the 13 scales. The group testing procedure dictated that the syllables be presented orthographically, but subjects were instructed to make their ratings, as much as possible, on the sound of the syllables, not on the basis of written form or articulation. The average rank score of the stimuli on the scales formed the basic measurement for the subsequent analysis (see Table 5.7).

TABLE 5.2

AVERAGE RANK SCORES ON RATING SCALES
EXPERIMENT IV

	pa	ma	ča	ba	ša	ṭa	na	za	la	ha	sa	da
hissy	3.7	2.8	7.7	3.3	9.4	5.1	3.5	9.1	3.1	4.9	10.4	3.2
vowel-like	5.8	7.3	4.6	6.0	6.0	6.0	8.0	6.8	8.5	7.5	5.8	5.6
bright	6.4	3.7	6.7	4.9	5.5	7.7	3.6	5.8	5.7	4.4	6.4	5.5
short	8.9	5.3	5.0	7.7	2.3	10.2	6.0	3.8	6.1	6.8	6.4	9.0
clear	7.0	5.0	5.2	6.5	2.9	8.3	4.9	4.2	5.4	4.0	5.7	7.3
harsh	6.5	3.7	10.2	6.7	6.1	7.9	5.1	8.5	3.8	5.8	6.3	7.4
distinct	6.0	4.9	5.3	5.8	3.7	8.4	4.9	5.0	5.2	4.7	5.0	6.8
high pitch	6.2	5.4	8.1	4.3	7.1	7.3	5.2	7.7	7.0	6.4	8.3	4.7
abrupt	7.7	3.2	7.9	6.3	3.4	8.6	4.1	4.4	3.6	5.2	3.8	7.7
even	6.0	6.3	2.8	6.7	4.6	6.3	6.0	4.0	6.7	5.2	5.0	6.6
loud	6.3	4.6	7.3	7.0	4.0	6.6	4.9	6.2	3.9	3.8	4.3	7.6
melodious	5.0	8.7	4.1	5.8	6.7	6.1	7.6	5.2	9.3	5.9	6.7	6.7
effortful	6.8	3.3	10.0	6.5	7.0	8.0	4.5	8.0	4.1	6.1	6.2	7.4

An indirect measure of similarity between any two sounds in the columns of Table 5.7 was obtained by matching their respective profiles of scores on the 13 scales. Pearson's product moment correlation coefficient was used as the index of profile similarity so that the matrix of similarity scores could be factor analysed.

Results:

Scale reliability estimates, expressed as estimated test - retest correlation coefficients were obtained by a method based on analysis of variance (Winer, 1971). Table 5.8 shows the estimated reliability of average rank scores for the phonemes on the scales.

TABLE 5.3

RANK ORDERED ESTIMATED SCALE RELIABILITIES
EXPERIMENT IV

	r
hissy	.98
short-long	.94
abrupt	.92
effortful	.91
harsh	.89
melodious	.85
loud	.83
clear	.81
pitch	.77
bright	.72
vowel-like	.69
even	.64
distinct	.64

When one asks which verbal scales do the subjects employ most consistently in rating the 12 stimuli, it is obvious from a rank ordering of the reliability estimates that they are just those which would be expected to discriminate well between the phoneme targets on the basis of the results of Experiments I and II. For example, pitch and loudness are very well established auditory perceptual qualities, but evidently play little role in differentiating between these 12 consonantal phonemes.

The correlation matrix of inter-phonemic similarity scores was factor analysed by the principle components method, which suggested that a two dimensional configuration was optimal for this set of data (for details see Appendix G). Simply from inspection of the obtained configuration (Figure 5.7) a strong correspendence with the results of

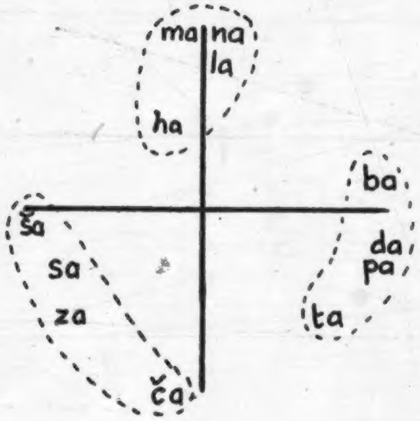


Fig. 5.7 Factor Analysis of Phone Rating Data Experiment IV

Experiments I and II is apparent. Rotation by the varimax criterion resulted in placement of the axes more or less where theoretical preference would have dictated. The two dimensional configuration accounted for 81.8% of the total variance. With the varimax rotation, 42.7% of the total variance was attributable to Factor I, the "duration" or "abruptness of onset" dimension and, 39.1% to the "sonority" or "resonance - hiss" dimension.

Furthermore, the same major three-way grouping by manner of articulation was obtained through cluster analysis as was found in the previous experiments.

CHAPTER VI

ANALYSIS AND DISCUSSION

In the previous chapter, based upon experiments involving similarity scaling of consonantal phones embedded in a CV syllabic frame, two hypothetical perceptual dimensions - a qualitative "resonance-hiss" dimension and a temporal factor, "duration," were identified. The perceptual configuration upon which this two-factor model was based was shown to be replicable, stable over different scaling methods, and substantially indifferent as to whether direct or indirect similarity rating techniques are employed to generate the matrix of perceptual proximities. A change in rank ordering within the sibilant cluster on the putative "temporal" dimension which occured when a change was made in the "carrier" vowel raised some question about the explanatory adequacy of this factor. The two-factor model yielded consistent predictions about the location in perceptual space of certain stimuli not included in the original scaling set, and thus could be said to meet minimal requirements of factor invariance.

However, while the proposed two-factor model may be theoretically attractive (see below) and consistent with the data to hand, it can lay no claims to uniqueness as a way of representing the perceptual relationships between the sounds included in the target set. This is evident when one considers the problems of dimensionality determination and rotation. The writer has relied rather heavily on the "subjective" criterion of theoretical preference to establish the dimensionality and orientation of the reference axes which will enable him to "adequately account for" the derived multidimensional perceptual configuration obtained through MDS of the input proximity matrix. In some instances the subjective criterion is clearly supported by the "objective" mathematical indices of the "most adequate" solution (e.g., Experiment IV). In other instances, particularly where the "objective" criterion fails, or none exists, the writer has of necessity exercised liberty in the choice of dimensionality and axis rotation.

Generally speaking, a better "fit" between the input proximities can be obtained by choosing a solution of greater dimensionality - or, in factor analytic terms, more of the common and total variance may be accounted for by extracting a higher number of factors. On the other hand, low dimensional solutions, and those factors which emerge first in a factor analysis, are more strongly determined, and hence more reliable than high dimensional scaling solutions and late emerging factors.

In the face of this problem of non-uniqueness, two general strategies appear to be open to the investigator. He may leave temporarily unresolved the question of rotation (and even that of dimensionality if he works simply with the raw proximity matrix) by inquiring what kinds of variables are most strongly associated with, or (in some experimental manipulation) most potently affect, the relative interpoint distances which determine the overall shape of the perceptual configuration. This strategy was also employed in the present study where multiple regression analysis with different kinds of variables (phonological, perceptual, and acoustic) was used to build predictive equations for the raw proximities and the derived (Kruskal, two-dimensional) interpoint distances. By examining the normalized predictor weightings for different multiple regression equations it should be possible to clarify the nature of the perceptual space.

The second strategy is to attempt to find independent empirical evidence for the factors isolated in the scaling study. There are a variety of ways this may be done. It may be possible to show, for example, that certain reliably ratable perceptual qualities can predict with a high degree of accuracy, scores on a particular interpretive factor. In this way a verbal characterization of the factor can be generated based on subjects' abilities to describe perceptual attributes of objects. If the physical correlates of some hypothetical perceptual dimension can be isolated and controlled (for example through synthesis in the case of speech perception), then the investigator is in a much stronger position to evaluate the perceptual reality and salience of his hypothetical factors. In this (ideal) situation, the experimenter would not only be able to predict the scaling configuration of a set of perceptual objects from a critical set of measurements of specific signal properties, but be able to manipulate the shape of the perceptual configuration by varying the relevant synthetic signal parameters.

Physical correlates of the two perceptual dimensions isolated in the scaling experiments were obtained (see below) by acoustic analysis of the /Ca/ stimulus set used in Experiment II. It was found that on the basis of these physical correlates, the derived perceptual configuration of the 12 sounds could be predicted with fair accuracy (Pearson's product moment correlation between the derived perceptual distances and interpoint distances predicted on the basis of the physical correlates = .81).

Regression Analyses of Interpoint Distances and Proximity Scores

Earlier in this paper (Chapters I and II), the possibility was entertained that features employed in phonemic recognition might be uniquely linguistic, rather than perceptual properties that mediate auditory recognition in general. If this is so, it may be anticipated that an "abstract" and singularly "phonological" distinctive feature schema such as that of Jakobson, Fant, and Halle or Chomsky and Halle should have considerable predictive power for the interpoint distances derived from MDS of perceptual proximities. Moreover, the raw proximity scores themselves should be predictable from a knowledge of the distinctive feature specification of the sounds.

On the other hand, results of the four MDS scaling experiments reported above suggest that the bulk of the systematic variation in the derived perceptual configuration is attributable to just two factors that seem to represent general auditory features of no unique linguistic character, and for which it is unnecessary to invoke any specific linguistic adaptation or specialization of the perceptual mechanisim. This, however, could be a mistaken impression, made possible by the non-uniqueness of the factor solution. A fair test of the utility of any proposed feature schema in predicting perceptual distances can be obtained through multiple regression analysis, if it can be assumed that the interpoint distance between two phonemic targets is a simple additive function of the number of contrasting feature values that serve to distinguish the two targets.

$$D_{jk} = r_1 |F_{1j} - F_{1k}| + r_2 |F_{2j} - F_{2k}| + \ldots + r_n | F_{nj} - F_{nk}|$$

where, D_{jk} = perceptual distance between targets j and k
F_{nj} = value of feature n for target j
r_n = normalized regression weight for feature n

Phonological feature systems do not make explicit claims for the relative importance of their component features, though some rank ordering of the features is often implied. The primary task of the regression analysis is to determine what the relative independent contribution of each feature in the system is for the particular perceptual configuration in question. The model is of course applicable with any a priori set of features, categorical or scalar. A stepwise multiple regression routine was used in the following analysis. The model can be expected to maximize the contribution of a few independent features to the prediction of the interpoint distances. The regression weights assigned to features in subsequent steps of the regression analysis will be highly influenced by the particular features extracted in previous steps of the analysis.

Subjects' responses to the /Ca/ set of stimuli in Experiment II were used in the following regression analyses. This set was chosen because it was based on a larger sample and the stimuli were better controlled than in experiment I. Five phonological feature systems were independently evaluated by the regression analyses with both the Kruskal two-dimensional interpoint distances, and the raw proximity scores as criterion variables. Feature System I was that of Jakobson, Fant, and Halle. System II was that of Chomsky and Halle. System III was the one used by Singh and Black (1968) which, for the present set of stimuli, is identical with the Miller and Nicely system except for the addition of a single feature (Liquid). The notable characteristic of System IV is that it incorporates three

TABLE 6.1

FEATURE SYSTEMS EMPLOYED IN REGRESSION ANALYSES

I Jakobson, Fant, & Halle	p	b	t	d	č	s	z	š	l	h	m	n
Vocalic	0	0	0	0	0	0	0	0	1	0	0	0
Consonantal	1	1	1	1	1	1	1	1	1	0	1	1
Compact	0	0	0	0	1	0	0	1	0	0	0	0
Grave	1	1	0	0	0	0	0	0	0	0	1	0
Nasal	0	0	0	0	0	0	0	0	0	0	1	1
Tense	1	0	1	0	1	1	0	1	0	1	0	0
Continuant	0	0	0	0	0	1	1	1	1	1	0	0
Strident	0	0	0	0	1	1	1	1	0	0	0	0

II Chomsky & Halle	p	b	t	d	č	s	z	š	l	h	m	n
Vocalic	0	0	0	0	0	0	0	0	1	0	0	0
Consonantal	1	1	1	1	1	1	1	1	1	0	1	1
High	0	0	0	0	1	0	0	1	0	0	0	0
Low	0	0	0	0	0	0	0	0	0	1	0	0
Anterior	1	1	1	1	0	1	1	0	1	0	1	1
Coronal	0	0	1	1	1	1	1	1	1	1	0	1
Voice	0	1	0	1	0	0	1	0	1	0	1	1
Continuant	0	0	0	0	0	1	1	1	1	1	0	0
Nasal	0	0	0	0	0	0	0	0	0	0	1	1
Strident	0	0	0	0	1	1	1	1	0	0	0	0

III Singh & Black	p	b	t	d	č	s	z	š	l	h	m	n
Voicing	0	1	0	1	0	0	1	0	1	0	1	1
Frication	0	0	0	0	1	1	1	1	0	0	0	0
Duration	0	0	0	0	0	1	1	1	0	0	0	0
Liquid	0	0	0	0	0	0	0	0	1	0	0	0
Place	1	1	2	2	3	2	2	3	2	4	1	2
Nasal	0	0	0	0	0	0	0	0	0	0	1	1

IV	p	b	t	d	č	s	z	š	l	h	m	n
Sibilant	0	0	0	0	1	1	1	1	0	0	0	0
Stop	1	1	1	1	0	0	0	0	0	0	0	0
Resonant	0	0	0	0	0	0	0	0	1	1	1	1
Voice	0	1	0	1	0	0	1	0	1	0	1	1
Place	1	1	2	2	3	2	2	3	2	4	1	2

V	p	b	t	d	č	s	z	š	l	h	m	n
Resonance	1	1	1	1	0	0	0	0	2	2	2	2
Duration	0	0	0	0	1	2	2	2	1	1	1	1
Voice	0	1	0	1	0	0	1	0	1	0	1	1
Place	1	1	2	2	3	2	2	3	2	4	1	2

binary manner features, one for each of the major perceptual clusters first observed in similarity rating data by Peters (1963) and subsequently found in Black (1968), Stevens and House (1971), as well as the data of the present experiments. System V contains two trinary features corresponding to the hypothetical two-factor basis for the derived perceptual configuration of the present study.

Results of the various stepwise multiple regression analyses are summarised in Figures 6.1a and 6.1b (for details of the analyses see Appen[illegible] H). The contribution of each feature in the final equation to the prediction of the criterion variable is indicated by the bar graph where the change in the squared multiple correlation coefficient which is associated with the feature in question is plotted on the y axis. This measure provides an estimate of the variance contribution of each feature in the final equation and may be expressed as a percentage. In interpreting these results, it should be remembered that the features are being applied only to a subset of the phonemic inventory and consequently some (such as Consonantal) do not have a chance to apply, while others (such as Strident) do not have their domain of reference adequately specified.

It is remarkable that virtually all of the predictive power of the Jakobson, Fant, and Halle, and the Chomsky and Halle feature systems can be accounted for by a single feature which opposes the sibilants to all the other consonants. This feature corresponds to the first factor

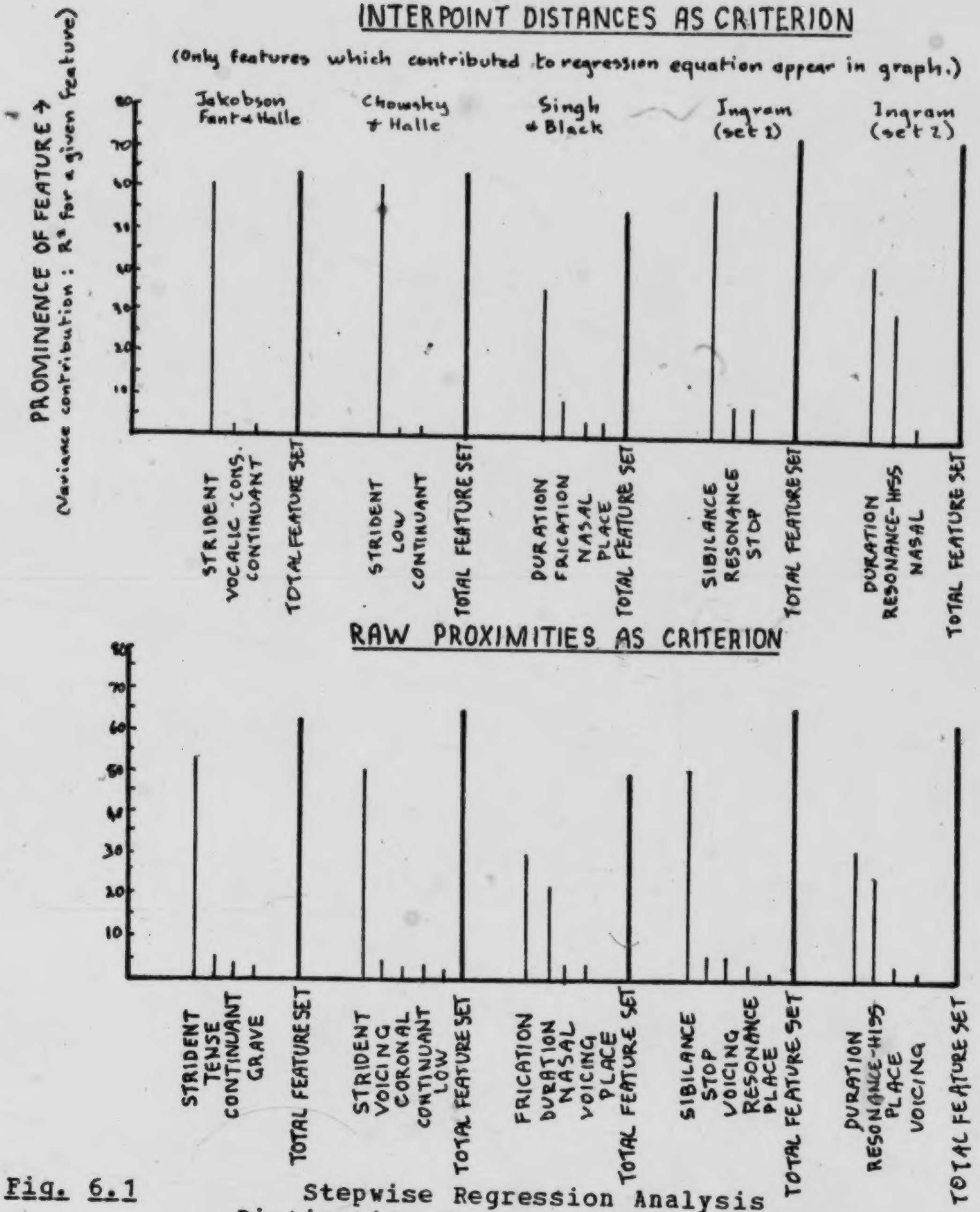


Fig. 6.1 Stepwise Regression Analysis
Distinctive Features as Predictors of
Derived Distances and Raw proximities

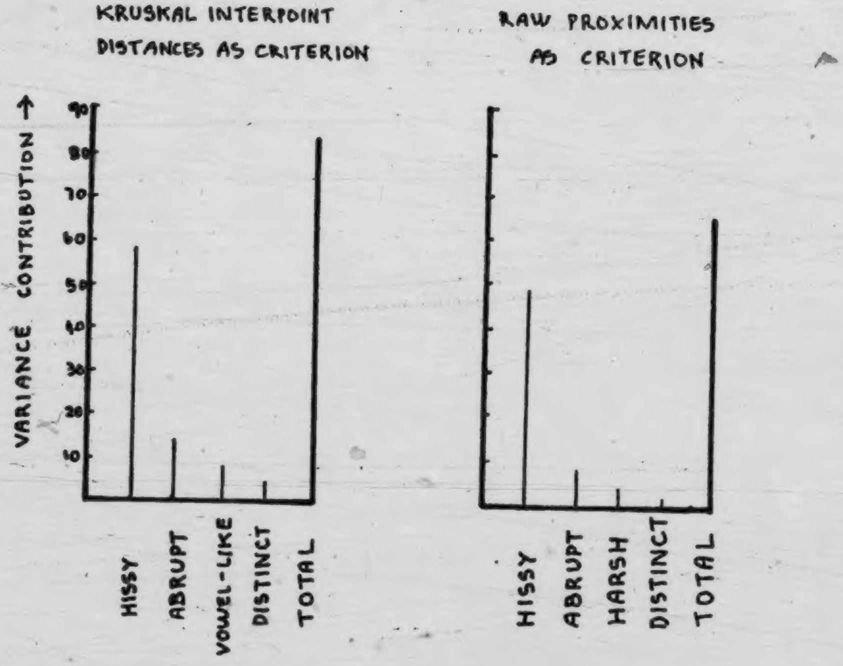
that was observed in the principal components analysis (pages 104-106 above). Essentially the same pattern is found in the analysis of both the interpoint distances and the proximity scores. No other feature accounts for more than 5% of the variance, which is roughly the cut-off level for deciding whether a feature makes a significant independent contribution to the prediction equation.

The Singh and Black (Miller and Nicely) feature system has two significant contributors to the prediction of the interpoint distances and the proximities. The relative importance of Duration and Frication for the distances is reversed for the proximities as criterion. Analysis IV once again shows the overwhelming importance of the sibilance contrast to both criterion variables. The other two "manner" features - Resonance and Stop make significant independent contributions to the interpoint distances, but Resonance gives way to Voicing in the prediction of the proximity scores. Feature System V, as would be expected, distributes the predictable variance more equally between the two trinary features of Duration and Resonance-Hiss. In terms of overall predictive power, there are no grounds for choosing between Feature Systems IV and V.

Generally speaking, the interpoint distances are more predictable than the raw proximity scores (R is , on the average, five percentage points higher for the interpoint distances.) This may be due to a certain "cleaning up" effect attributable to the scaling algorithm where some of the error present in the raw proximity scores is corrected for when each interpoint distance is determined as a function of all the other interpoint distances in the configuration. However, a certain amount of information loss also seems to occur when the proximities are transformed into distances in a two dimensional space. The Voicing feature appears in all five prediction equations when the proximity scores constitute the criterion but in none of the corresponding equations for the interpoint distances.

Perceptual Rating Scales as Prediction Variables

The somewhat higher predictability of the interpoint distances is also indicated when the perceptual rating scales of Experiment IV are used as predictors of the interpoint distances and raw proximity scores of Experiment II. The results of this regression analysis are summarized in Figure 6.2 (details in Appendix I). The table of average rank scores (Table 5.7) collected for Experiment IV served as the basis for correlating rating scale scores with interpoint distances and proximitiy scores. Ideally, the actual stimuli used in Experiment II should have been used rather than the rankings obtained in Experiment IV under non-auditory stimulus presentation. However, the latter data were readily at hand and while the final level of prediction might have been somewhat higher if the stimuli of Experiment II had been auditorily presented while subjects made their ratings, it is doubtful that the *pattern* of regression weightings would have been significantly different.


KRUSKAL INTERPOINT DISTANCES AS CRITERION

RAW PROXIMITIES AS CRITERION

Fig. 6.2 Stepwise Regression Analysis
13 Perceptual Scales as Predictors
Derived Distances and Raw Proximities

The "Hissy" scale dominates the prediction equation as might be expected of a dimension that clearly and reliably distinguishes the sibilant consonants from the rest. The "Abrupt" scale is second in importance for both the distances and the proximities. "Vowel-like" appears to contribute significantly to the prediction of the interpoint distances but not the proximity scores - just as the binary Resonance feature in regression analysis IV in Figure 6.1 above seems to be of slightly less importance for the proximities than the distances.

The important question to be answered with the help of these regression analyses is what impact do they have on the tenability of the two-factor model proposed earlier on the basis of the MDS studies? The implications appear to be the same as those of the principal components analysis (Figure 5.7). At a cost of sacrificing 10 - 18% of the predictable variation in the interpoint distances (4 - 8% in the case of the proximities), the temporal dimension may be abandoned and the bipolar resonance-hiss dimension collapsed into a monopolar hiss-nonhiss or Sibilance factor. Acceptance of this option would simplify the factor structure even further, but it seems to throw away important information contained in the perceptual structure. However, the strongest grounds for reluctance to accept the single factor solution stems from analysis of the acoustic correlates of the two hypothesised factors of Duration and Resonance-Hiss reported in the following section.

Physical Correlates of the Hypothesised Perceptual Dimensions

The approach that was used with the phonological feature systems and the perceptual rating scales, of attempting to predict interpoint distances or raw proximities from sets of possibly relevant predictor variables, was not employed in the case of the physical correlates of the perceptual configuration. It was felt that the problem of sampling from the domain of possibly relevant acoustic variables was too formidable. In certain restricted areas, such as the perception of steady state vowels (Pols, 1969) where all relevant information must be restricted to the spectral domain, it is feasible to think in terms of obtaining an unbiased and exhaustive sample of the total "acoustic space". However, with the time-varying spectral functions of consonantal sounds, the problem of unbiased sampling seems to be too open ended. On the other hand, the hypothetical dimensions extracted from the MDS analyses did suggest specific characteristics of the signals that subjects appeared to be using in making their similarity judgements. Attention was therefore focused upon physical correlates of the hypothesised perceptual dimensions rather than the (uninterpreted) interpoint distances.

As a starting point for the acoustic analysis, broad band spectrograms (b.w.=300 Hz; range = .016 - 16kHz; Kay Electro-Sonagraph) and high temporal resolution oscillograms (via computer controlled read out of the digitalized signals; see Roszypal, 1973) were made of the 12 /Ca/ and the 12 /Ci/ post digitalized stimuli used in Experiment II (see Appendix J). The /Ca/ stimulus set, which had the clearest perceptual structure, was chosen for detailed measurement and analysis. The axes of the Kruskal two-dimensional configuration for the /Ca/ set were graphically rotated (preserving orthogonality) to what appeared to be the theoretically most satisfying orientation, and the loadings of each stimulus on the rotated dimensions were recorded (see Table 6.2).

TABLE 6.2

STIMULUS LOADINGS ON ROTATED KRUSKAL DIMENSIONS

STIMULUS	LOADING RES-HISS	 DURATION
sa	[illegible].66	1.02
pa	0.50	-0.95
ča	-1.19	0.01
ma	0.81	-0.19
ta	-0.59	-0.95
la	0.97	0.08
da	-0.21	-0.74
ha	0.54	0.05
ša	-0.94	0.66
za	-0.49	1.21
na	0.84	0.22
ba	0.37	-0.80

There was no difficulty finding acoustic correlates of the temporal dimension. It correlated .95 with the physical duration of the consonantal portion of the syllable and .94 with the duration of the whole syllable. (Segmentation was not difficult; see Appendix J for the measurements.) It would be unreasonable to expect results clearer than these.

The Resonance-Hiss dimension, however, posed some difficulties. It may be grossly, but rather inaccurately, characterised by a separation of high and low frequency spectral energy bands. All the resonants (with the notable exception of /h/) have a low frequency, periodic glottal energy source. Sounds at the other end of this dimension are characterised by relatively high frequency spectral energy. However, the *quality* or type of energy present appears to be relevant and not merely its locus in the frequency domain. The syllables /la/ and /ča/ take extreme values on the Resonant-Hiss dimension yet both have substantial energy concentrations in the 3-5 kHz band. The crucial difference would seem to be that in one case the spectral energy distribution is highly organized in terms of harmonic and formant structure, but in the other it is random, or lacking in any spectral organization.

The basic problem in obtaining some satisfactory acoustic correlate of the Resonance-Hiss dimension resided in the fact that currently available acoustic analysing devices are ill suited to detecting such a distinction, though the human ear and other biological sound analysing devices (Suga, 1972) apparently are not. Instrumental limitations therefore led to a choice, as the best practical approximation to an adequate physical characterisation of this dimension, of a simple bandfilter function that optimally predicted the Resonance-Hiss factor loadings.

The Resonance (or, more accurately, the Vocalic) and the Hiss components of the stimuli were extracted separately by simultaneous low-pass (LP) and band-pass (BP) filtering of the post-digitalized stimuli. Optimal filter settings (LP < 200 Hz, -48dB/octave, Rockland Programmable Filter series 1520; Bp = 2.7-5.6 kHz, -32dB/octave, Audio Frequency Filter type 400) that would maximally differentiate the stimuli on the Resonance-Hiss dimension loadings were determined on the basis of spectrographic analysis and trial and error. The output of each spectral band filter was fed into a dual channel Frokjar-Jennson Intensity Meter, operating on an integration time base of 20 msec. This provided a smooth but sufficiently time sensitive intensity trace which was recorded on an Elma-Schonander four channel Mingograph at an operating speed of 100 mm/sec. The intensity meter provides for either a linear or a logarithmic scale for registering intensity over an operating range of 50dB. The logarithmic setting has the effect of magnifying differences at the low intensity levels of registration at the expense of differences at higher intensity levels. For purposes of clear segmentation into consonantal and vowel portions of the syllables, a Duplex Oscillogram was also obtained. For the instrumental configuration see Figure 6.3 below.

It was hypothesised that the stimulus loadings on the Resonance-Hiss dimension could be adequately approximated by some linear combination of the low and high frequency band output levels where the weightings of the two bands will be opposite in sign (see following equation):

$$\text{PRES} = a(\text{LP output}) - b(\text{BP output}) + C$$

where, PRES = least squares match with loadings on Resonant-Hiss dimension
a = low frequency band output weighting
b = high frequency band output weighting
C = some arbitrary (uninterpreted) constant.

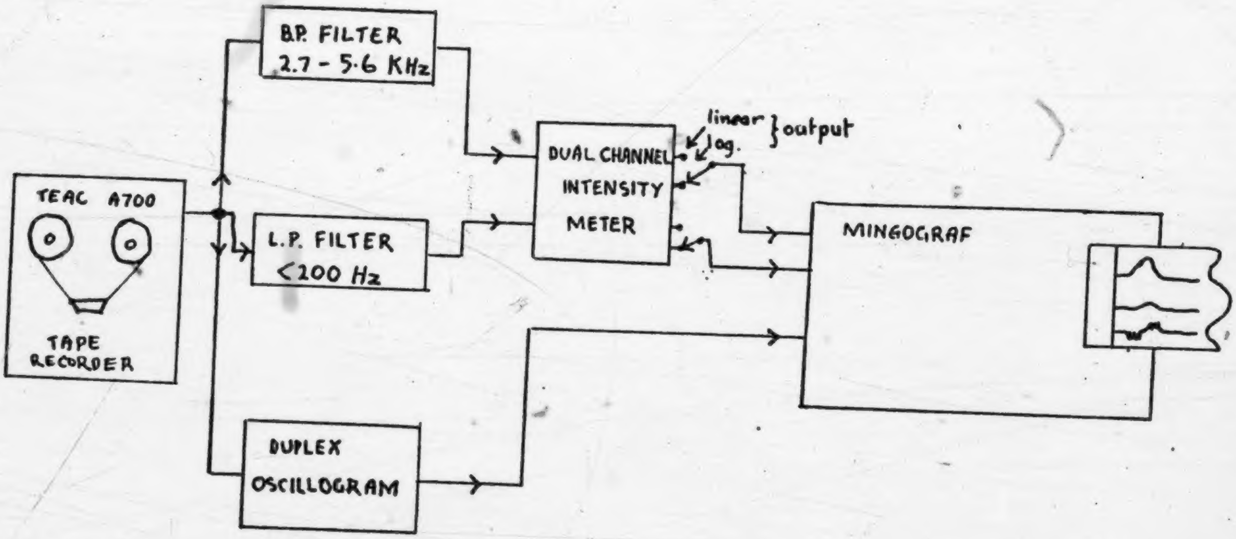

B.P. FILTER
2.7 - 5.6 KHz
linear
log.
output
DUAL CHANNEL
INTENSITY
METER
TEAC A700
TAPE RECORDER
L.P. FILTER
<200 Hz
MINGOGRAF
DUPLEX
OSCILLOGRAM

Fig. 6.3 Instrumentation for Acoustic Analysis of "Resonant-Hiss" Dimension

Regression analysis was used to determine the regression weights and constant in this equation which would optimally predict the Resonance-Hiss loadings. Four sets of predictor variables were tried based upon the linear or the logarithmic intensity meter scale outputs and whether or not the area function or the peak amplitude of the intensity traces for the consonantal portion of the syllable was measured. Results for the four sets of predictor variables did not differ greatly (see Table 6.3, Appendix k for details). The peak amplitude measurements resulted in slightly better predictions than the area function (total energy) measurements, and the log scale fared somewhat better than the linear scale. The normalized regression equation for the prediction of the Resonance-Hiss dimension loadings in terms of the band-pass, log.scale, peak

TABLE 6.3

PHYSICAL PREDICTORS OF "RESONANCE-HISS" DIMENSION LOADINGS

MEASUREMENT SCALE	MULTIPLE REGRESSION COEFICIENT
AREA FUNCTION LIN.	.84
AREA FUNCTION LOG.	.85
PEAK AMPLITUDE LIN.	.89
PEAK AMPLITUDE LOG.	.92

amplitude (BPGP), and the low-pass, log.scale, peak amplitude (LPGP) was:

$$PRES = .33(LPGP) - .75(BPGP)$$

Note: This regression equation was derived from measurements of the oscillograph trace deflections from the baseline. The units are therefore arbitrary. They may be converted into dB ratings by means of the calibration curves for the two traces given in Appendix L.

As the above equation indicates, the peak intensity level of the high frequency band is considerably more important for the prediction of the factor loadings than the low band peak.

Although the Resonance-Hiss dimension loadings may be quite successfully predicted by the simple bandfilter function developed in this study, there is some doubt that this perceptual continuum is correctly characterized in terms of an energy by frequency-band analysis. For example, the bandfilter analysis fails to predict the high loading of /la/ on the "resonant" pole of the Resonance-Hiss continuum because it takes no account of the kind of energy present in the 3-5 kHz band. It fails to distinguish spectrally coherent signals from those that lack organization in the frequency domain. Such organization could be due to the nature of the source signal (the spectral coherence of the glottal tone), or the "shaping" function of a resonator, or, conceivably, to both of these factors.

With this in mind, as an alternative to using the low frequency voicing component (LPGP) to capture the Resonance pole of the Resonance-Hiss continuum, an attempt was made to quantify the notion of spectral organization, or "formant structure". Two trained phoneticians, experienced in spectrographic analysis of speech were asked to rate the spectrograms of the 12 /Ca/ stimuli for the presence of "formant structure" on a four point scale ranging from "strongly apparent" to "no detectable formant structure". The two sets of independent ratings (inter-rater reliability estimated at rho = .86) were averaged and entered into a regression equation with the other predictor variable, the high-band peak amplitude, BPGP. This resulted in a multiple regresion coeffeiient of R = .92 for the Resonance-Hiss loadings (see Appendix K for details). Table 6.4 shows that the low-band peak amplitude (LPGP) and the average formant structure ratings (FRAV) are significantly intercorrelated, but both of these variables are sufficiently independent of the high-band peak amplitude (BPGP) to justify combining either FRAV or LPGP with BPGP to define a single, bipolar, Resonance-Hiss dimension. The notion of spectral coherence seems to play a significant role in the prediction of the

TABLE 6.4

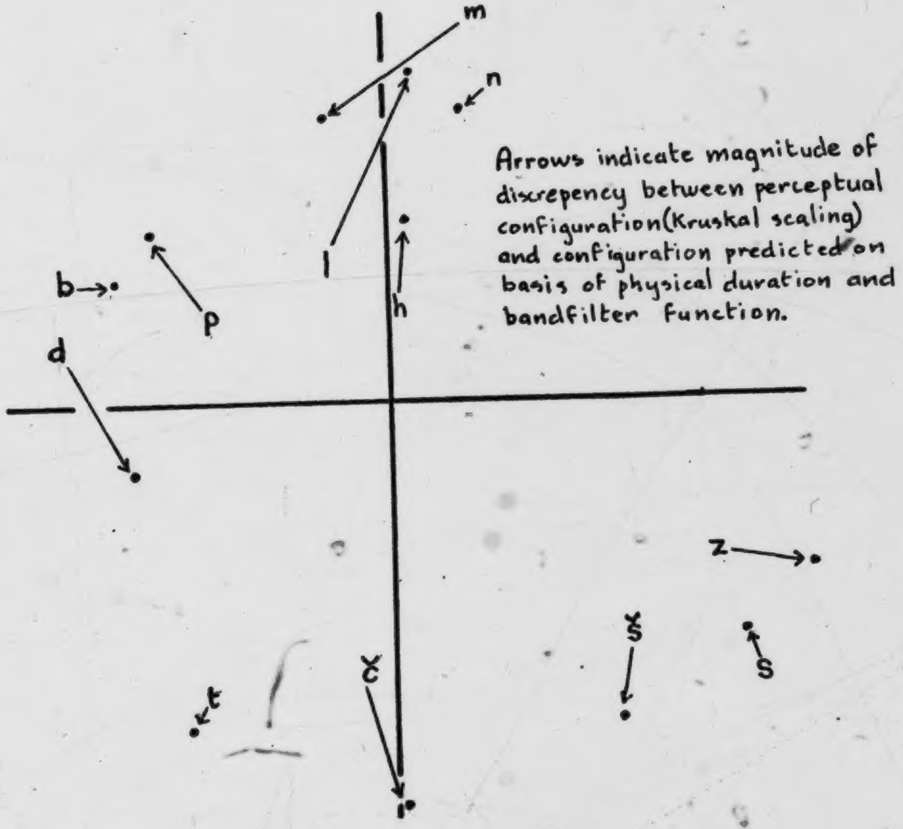
CORRELATIONS BETWEEN PREDICTOR VARIABLES

		BPGP	LPGP
HIGH BAND PEAK AMPLITUDE	BPGP		
LOW BAND PEAK AMPLITUDE	LPGP	-.306	
MEAN FORMANT STRUCTURE RATING	FRAV	-.306	.623

Resonant-Hiss dimension and, by implication, in the determination of perceptual structure. But because it could not be fully operationalized (i.e. instrumentally measured) this variable was not used as a physical predictor in the reconstruction of the perceptual configuration.

Figure 6.4 indicates how well the two-dimensional Kruskal configuration for the stimuli may be predicted from the physical correlates of the two hypothetical perceptual dimensions - the bandfilter regression function, and the physical duration of the consonantal portion of the syllable. As previously mentioned, the two sets of interpoint distances in Figure 6.4 are quite highly correlated (r= .81).

A recent experiment by Pols (1974) on the physical correlates of Dutch CVC syllables provides an interesting corroboration of the _physical_ Resonance-Hiss dimension isolated in this study. He bandfiltered 270 spoken CVC syllables with 17 parallel filters whose bandwidths approximately matched the frequency resolution of the human ear. The output intensity of each filter was sampled every



Arrows indicate magnitude of discrepency between perceptual configuration(Kruskal scaling) and configuration predicted on basis of physical duration and bandfilter function.

Fig. 6.4 Reconstruction of two-dimensional perceptual configuration on basis of physical predictor variables

10 msec. Each of the 14,111 resulting 10 msec. samples may be described as a point in a 17 dimensional space, having co-ordinate values equal to the levels of the 17 filters. These variables were subjected to a Principal Components analysis. The first factor which emerged (accounting for 55.1% of the total variance) was "a very efficient discriminator between sonorant and non-sonorant sounds." The nature of this factor may be illustrated by the graph of the first eigenvector in Figure 6.5. Peak values in the eigenvector closely correspond with the optimal centre frequencies in the high and low frequency bands that were

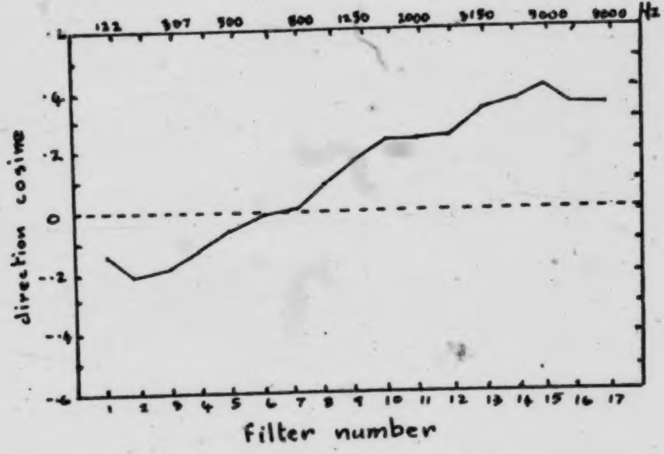


Fig. 6.5 First eigenvector of the variance-covariance matrix of bandfilter spectra (from Pols, 1974)

used to predict the Resonant-Hiss loadings in the present study. Unlike the present study, Pols' speech samples were drawn from vowel as well as consonantal segments of syllables and this suggests a perceptual role for the Resonance-Hiss dimension that could only be a speculative extrapolation from the present data, but which Pols explicitly demonstrated by example (see Pols, 1974, p.90). Specifically, the Resonant-Hiss dimension proved to be a highly effective basis for vowel - consonant segmentation, which is often recognized in such diverse areas as automatic speech recognition and phonological theory as the most fundamental distinction in the segmental analysis of speech.

Although the Resonant-Hiss dimension emerges strongly in both the scaling solutions and the regression analyses in the present data, the status of the temporal dimension is more equivocal. It appears to account for a good deal less of the variance in the perceptual configuration and is not stable over a change in the "carrier" vowel.

To show that the transposition of the sibilants on the second dimension of the perceptual configuratiion (Figure 5.4) is consistent with the interpretation that it represents a "consonantal duration" or "abruptness of syllable onset" factor, it will be necessary to show that this change is a perceptually real phenomenon. A verbal rating scale experiment, similar to Experiment IV of the present study, but with /Ci/ syllables, could provide the relevant perceptual evidence. Measurements on the physical duration of the consonantal segments of the /Ci/ stimuli showed that /č/ was considerably longer in this than the /Ca/ set. However the correlation between the physical duration and loadings on the temporal dimension of the Kruskal scaling solution dropped from .95 for the /Ca/ set to .75 for the /Ci/ set.

One hypothesis that could explain the perturbation in the perceptual configuration with the change in carrier vowel is that the perceptual prominence of certain auditory features relevant to the recognition of the consonant are subject to differential backward masking effects by different vowels. This hypothesis gains plausibility when coupled with the suggestion that the two-dimensional representation favoured by the analysis thus far may be an over-reduction of the proximity matrix. In other words, features that went undetected by the MDS analysis may have been differentially enhanced or suppressed in interaction with the carrier vowel. In response to this suggestion it can only be observed that for reasons of mathematical determinancy and replicability of findings with MDS, it is advisable to keep the dimensionality of the solution low. In this way one may be assured of capturing at least the most important of the feature dimensions contained in the proximity matrix.

Finally, the possibility cannot be dismissed that the problematical loadings of the sibilants on the "temporal" dimension for the /Ci/ set is an experimental artifact. It is apparent from the oscillograph tracings (Appendix J) that there is some temporal clipping of the carrier vowel in the longest syllables of the /Ci/ set. This occured because of time sample limitations in the gating program used for digitalized storage of the stimuli. The clipping was noticed at the time the stimuli were constructed but because it was barely detectable in playback, it was not judged to be a potentially significant influence on the subjects' ratings. In retrospect this may have been a mistake. In any event, it clouds a potentially important point in the analysis of the data of Experiment II.

Broader Discussion of Findings

Quite clearly it would be simplistic to suggest that the two perceptual dimensions identified in these experiments comprise the necessary and sufficient set of auditory features that listeners' employ for consonantal phonemic recognition. Perhaps the most obvious objection to this interpretation of the results of the present study is that the range of stimuli is too restrictive. The use of only 12 stimuli per scaling set quite severely restricts the upper limit on the number of readily interpretable and reliable dimensions that is likely obtainable. On the other hand, an effort was made to broadly sample from the domain of auditory variability manifest in the consonantal sounds of English, so that those dimensions which are obtained should be the major ones and demonstrable in larger studies which more adequately represent that set. In this connection, the similarity rating studies of Black (1968) and Singh, Woods, and Becker (1972) are important because of the large number of stimuli they employed.

The plot of the first two principle components of Black's solution (Figure 3.5) shows that (within rotational invariance) his data agree quite well with the two dimensions isolated in the present study. Singh et al'.s (1972) results (Figure 3.6) are more problematical. The ABX condition, which comes closest to the triadic comparisons scaling technique used in the present experiments, yields a two-dimensional configuration that generally matches expectations of the two-factor model. Clearly though, there is poor agreement between the Duration and Resonance-Hiss model and the results of Magnitude Estimation (ME) and seven-point (SP) scaling. Lack of raw data for the other reported studies of consonantal similarity rating (Peters, 1963; Pruzansky, 1970) makes comparisons more difficult. But from the authors' own reports (Chapter III, pages 56 and 61), it seems that there is substantial agreement with the findings of the present study in obtaining a perceptual configuration where the consonants cluster by traditional "manner of articulation" groupings in a space definable by two orthogonal factors of sound duration and quality.

In assessing the theoretical significance of the present findings, it may be crucially important to note the agreement between Graham and House's (1970) data (see Figure 3.2) on perceptual confusions of young children and the perceptual structure predicted by the two-factor model. Admittedly, there are some discrepencies in the obtained perceptual configuration, but considering the nature of the data (and the fairly high "stress" rating) this is only to be expected. The Graham and House experiment is one of very few reported studies on the development of perceptual capabilities for speech recognition at the phonological level. (Most developmental investigations - probably for methodological reasons - have concentrated upon the acquisition of phonological contrasts in speech production.)

Developmental data are, however, of vital theoretical interest because the order of acquisition of different phonemic contrasts can potentially provide important information about the perceptual processes underlying phonemic recognition in the "linguistically competent" adult. One may reasonably hypothesise that those phonemic contrasts which are mastered at a very early age, correspond with auditory distinctions that are most "natural" for the perceptual apparatus - discriminations made with high reliability without need of extensive "ear training". On the other hand, late emerging phonemic contrasts (of which Voicing has been claimed to be one of the last: Shvachkin, 1948, in Fergurson et al., 1973; also, Garnica, 1971) would presumably constitute "difficult cases" for the perceptual apparatus, requiring perhaps highly complex perceptual processing of the signal beyond some "primary auditory" level of neural representation. If this developmental hypothesis is correct, then the spatial structure of confusion matrices obtained from subjects whose perceptual capabilities for phonemic recognition are incompletely developed, should largely reflect those primary auditory dimensions that are most salient to the relatively "language naive" ear. Thus arguably, the agreement between the Graham and House data and the findings of the present study supports the hypothesis that the most important determinants of the perceptual configuration obtained in the present study are not specifically linguistic distinctive features, but perceptual dimensions that may subserve auditory recognition in general (Hypothesis II in Chapter i). This conclusion is also indicated by the notable failure of "abstract" phonological feature systems to contribute significantly to the prediction of the interpoint distances or the proximity scores - beyond what might be anticipated from the two-factor model.

It may be argued that the experimental results are not counterindicative of the existence of specifically linguistic feature detection in phonemic perception, but merely that such features fail to show up in the MDS of similarity judgements. It does in fact seem that a rather low order of perceptual processing is being tapped by these experiments. Literally interpreted, the two-factor model may be characterised as a device which reliably segments the acoustic signal into broad perceptual categories - sufficient to distinguish vocalic from consonantal segments, and within the latter class, to differentiate "manner of articulation" groupings - Stops, Resonants, Sibilants, (Soft Fricatives?). These groupings roughly suggest the probable limits on the resolving power of the simple two-factor model. More complex perceptual decoding would seem to be required to attain the level of phonemic and phonetic resolution characteristic of the "perceptual competence" of the native listener.

It is not unreasonable to suggest that a good deal of perceptual learning (specific linguistic training of the auditory perceptual apparatus) is required before listeners can readily differentiate phonemic targets within the major manner groupings yielded by these experiments. It is well known, for example, that the stop consonants require for their mutual differentiation acoustic cues such as formant transitions which are context dependent and therefore require a complex mapping between acoustic signal and perceptual target that some writers (Liberman et al., 1967) have labeled "encoded". These same kinds of cues apparently play a significant, though progressively less important role for mutual discrimination within the resonant and fricative consonantal sub-groups. Correspondingly, dichotic listening studies show a significant but progressively decreasing right ear (left hemisphere) effect for natural (or synthetic) stop consonants (Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970) resonants (Haggard, 1971), and fricatives (Darwin, 1971).

Interestigly, no dichotic listening studies have reported lateralization effects for selected sets of stimuli drawn from across rather than within the perceptual groupings found in the present experiments. It would be predicted that for such stimulus sets no significant lateralization effect would be found. Even within these perceptual clusters, the relative strength of the lateralization effect for particular phonemic targets may be predictable. Note for instance that /t/ is separated quite distinctly from the other stops /p,b,d/ on the Resonant-Hiss dimension, presumably by its stronger "hard aspiration" (see Figures 5.3, 5.4). It may therefore be expected to yield a correspondingly weaker lateralization effect than the other stop consonants. Studdert-Kennedy and Shankweiler (1970) found in fact that of the six stop consonants /p,b,t,d,k,g/, /t/ ranked lowest in terms of lateralization effect under dichotic listening. These results do not imply that the left hemisphere is uniquely specialized for the detection of certain kinds of phonetic or phonemic features.

Recent experiments (Carmon and Natchson, 1973; Papcun, Krashen, Terbeek, Remington, and Harshman, 1974) have obtained right ear superiority for the dichotic perception of clearly non-speech stimuli and, on balance, the evidence suggests that a general facility with the extraction of temporal sequencing may be important for explaining the lateralization of certain kinds of speech sounds, rather than some speech or language specific perceptual capability. Whatever the nature of the relevant differential hemispheric capability may be, the fact that a perceptual learning factor is important is strongly indicated. (Compare the performance of novice vs. experienced Morse code operators in the dichotic perception of Morse code signals in Papcun et al., 1974.)

In short, the "distinctive feature" contrasts which produce lateralization effects in dichotic listening appear to be those for which the auditory system requires special adaptation, which is presumably obtained through perceptual learning at some early stage of language acquisition. These, however, are not the prominant perceptual contrasts that emerged in this study. On the contrary, their salience was weak to the point of undetectability by the analytical methods employed in this study.

CHAPTER VII

SUMMARY AND SOME SUGGESTIONS FOR FURTHER RESEARCH

The experiments reported in this paper, and a review of the relevant scaling literature, suggest that a small number of perceptual dimensions (two or three) are of paramount significance for the recognition of consonantal sounds embedded in an isolated monosyllabic frame. The strongest and most easily replicable dimension, which was labelled Resonance-Hiss, is conceivably employed, not simply for broadly differentiating the consonants as perceptual targets, but for providing the acoustic basis for segmentation of the signal into consonantal and vocalic frames - an operation which would likely provide essential information for the functioning of higher-order stages of perceptual-linguistic processing. Although the loadings of the sounds on the Resonant-Hiss axis of the MDS solution were fairly accurately predictable on the basis of a simple bandfilter analysis of the stimuli, it would seem to be an oversimpification to characterise this dimension, in acoustic terms, as a "low-to-high" frequency continuum. The notion of "degree of spectral coherence" was introduced to make provision for the role that the resonating cavity, coupled with the source signal, appears to play in locating sounds along this continuum.

A second dimension identified as a temporal factor was found to be replicable in the present experiments, and was discernible also in other studies employing similarity scaling (Peters, 1963; Black, 1968; Pruzansky, 1971; Singh, Woods, & Becker, 1972) as well as a study of perceptual confusions of young children obtained under non-noisy listening conditions (Graham & House, 1971). Loadings on this dimension correlated highly with the duration of the consonantal segments of the test syllables.

A third, Voicing dimension, was apparent in the results of Experiment I, but was not clearly discernible in subsequent experiments. It has often been claimed, on the basis of experiments with white-noise masking, that Voicing is the most salient distinctive feature contrast amongst the consonants. However, in noting the strength of the Voicing dimension commentators (such as Shepard, 1972; or Studdert-Kennedy & Shankweiler, 1970) have tended to overlook the specific experimental conditions that resulted in the preservation of auditory information in only the lowest frequency band. Similarity rating studies, in the absence of high frequency masking, suggest that Voicing is not a strong perceptual dimension for phonetically untrained listeners.

This finding is of general interest for the study of perceptual processes underlying speech recognition at the phonological level because it points to an imbalance of focus in current theoretical discussions which this paper may help to redress. A great deal of research effort has been directed to the study of phonetic contrasts such as Voicing and Place of Articulation for which the recognition problem stated in terms of a mapping between reliable acoustic cues in the signal and the invariant perceptual target, is known to be quite complex. It is a moot point whether these kinds of fine perceptual discriminations require the postulation of a genetically "pre-wired" phonetic feature detection capability of the human brain (as argued in a recent review article by Cutting and Eimas, 1974). This author is inclined to the view that current evidence for such a "nativist" position is highly equivocal and that the adaptability of the auditory-perceptual system, in conjunction with a learning process directed by the phonological exigencies of the hearer's native language, provides a sufficient schema for the experimental data presently at hand. Resolution of this debate will only be possible when a great deal more reliable information is obtained about the developmental timetable for the acquisition of linguistically relevant sound contrasts, and when basic processes of auditory discrimination and recognition are better understood than at the present time.

Setting aside the complexities of the "nativist-empiricist" debate in relation to speech perception, there remains the broader and feasibly-answerable question of whether a specifically "phonetic" or "phonological" level of perceptual processing is clearly discernible in "phonemic recognition", as the term has been used in this report. The alternative view, which appears to be favoured by the results of the present experiments, is that subjects' responses (as manifest, for example, in similarity judgements to simple CV stimuli) are most readily explicable, not in terms of the posession of a set of specifically linguistic feature detectors, but in terms of features that reflect plausible response parameters of mammalian auditory systems in general, and the human auditory system in particular. The Resonant-Hiss factor which appeared to be the most important determinant of the derived perceptual configurations in the present experiments is, arguably, not a specifically linguistic dimension, but a general auditory continuum that subjectively separates "resonant", "musical", and "pleasant" sounds from those that are "noisy", "harsh", and "unpleasant". Dimensions of auditory discrimination (such as Place of A[illegible]ulation, or Voicing, particularly in stop consonants) for which special linguistic adaptation of the auditory mechanism seems to be necessary (either through learning or heredity, or both) appear to play a secondary role in phonemic recognition.

The imbalance referred to earlier, which lays undue stress upon the unique character of speech recognition *vis a vis* other forms of auditory perception, would seem to result from an over-concentration of research attention upon a highly restricted segment of the auditory domain that is generally utilized by listeners in phonemic recognition. Of course, the question of phonological factors in phonemic recognition has only just begun to be raised in an interesting way by experimental studies of speech perception and the methodology of MDS has yet to be fully exploited. Terbeek and Harshman (1971) have offered some highly suggestive evidence in a cross-language study of vowel perception, that language-specific, phonological factors, play an important role in the structure of the perceptual space. An investigation, similar to theirs, of consonantal perception should prove interesting.

In the course of the present investigation a preliminary but unsuccessful attempt was made to test the "speech-mode" hypothesis (Liberman et al., 1967) in the context of the MDS paradigm. It was hoped that by gating out the steady state vocalic portions of the CV stimuli, and replacing them with a periodic, synthetic, "buzz" of roughly the same fundamental frequency, intensity, and duration as the replaced vowel, it would be possible to generate a set of stimuli that subjects would hear as "non-speech" sounds, yet with the essential acoustic cues for the recognition of the initial consonant preserved. Unfortunately, the phonetically untrained ear is not so easily fooled by such acoustic conjury (for which Roszypal's elegant PDP-12 gating program is in no way to blame). With repeated stimulus presentations that are necessitated by the Triadic Comparisons method, most of the supposedly "non-speech" stimuli became readily recognizable, "funny speech" sounds produced, as one subject put it, by "some sleepy dragon".

A more promising approach (which time, and some technical problems with the speech synthesiser prevented the writer from exploring sufficiently for this report) to testing, more strongly, the validity of the interpretation assigned to the MDS results of the present study, involves the use of purely synthetic stimuli. If the two major acoustic parameters isolated in these experiments are in fact the variables largely responsible for the shape of the derived MDS configuration, then, degrading the original set of stimuli in such a way as to preserve only the variation on these two acoustic variables should not substantially alter the derived perceptual configuration.

To preserve the temporal dimension, (Consonantal Duration or Abruptness of Syllable Onset) the temporal envelope of the stimulus is required. For variation in the Resonant-Hiss dimension, the bandfilter intensity functions (one for the high and the other for the low frequency band) that were obtained from the acoustic analysis of the original set of scaling stimuli (see Figure 6.3) may be used to control the Hiss Amplitude and the Voice Amplitude parameters of the PAT speech synthesiser. In this manner a "reconstituted" set of scaling stimuli could be obtained that would match the original scaling set of CV syllables just with respect to those acoustic variables thought to be responsible for the basic shape of the derived perceptual configuration. MDS of this "new" set of stimuli should yield a configuration in substantial agreement with the old - if the hypothesised basis for the subjects' similarity judgements is correct.

More generally, MDS experiments with synthetic auditory stimuli are needed to test the validity of some of the parametric assumptions of the MDS model itself in the context of auditory perception. Until this is done, the potential utility of MDS to problems of speech and auditory perception will remain in some doubt.

REFERENCES

Attneave, F. Dimensions of similarity. American Journal of Psychology, 1950, 63, 516-556.

Black, J. W. Interconsonantal differences. Archivio di Psicologia, Neurologia, e Psichiatria, 1968, 29(3), 277-293.

Bricker, P. D., Pruzansky, S., & McDermott, B. J. Recoverability of spatial information from subjects clusterings of auditory stimuli. Paper presented at the meeting of The Psychonomic Society, St. Louis, October, 1968.

Carroll, J. D., & Chang, J. J. An analysis of individual differences in multidimensional scaling via an N way generalization of "Ekart-Young" decomposition. Psychometrika, 1970, 35, 283-319.

Chomsky, N. & Halle, M. The Sound Pattern of English. New York: Harper and Row, 1968.

Chomsky, N., & Halle, M. Some controversial issues in phonological theory. Jorurnal of Linguistics, 1965, 1, 97-138.

Conteras, H. Simplicity, descriptive adequacy, and binary features. Language, 1969, 45(1), 1-8.

Corcoran, D. W. J., Dorfman, D. D., & Weening, D. L. Perceptual independence in the perception of speech. Quarterly Journal of Experimental Psychology, 1968, 20, 336-350.

Cutting, J. E., & Eimas, P. D. Phonetic feature analysers and the processing of speech in infants. Speech Research - Haskins Laboratories status report, 1974, 37/38,45-64.

Darwin, C. J. Ear differences in the recall of fricatives and vowels. Quarterly Journal of Experimental Psychology , 1971, 23, 386-392.

Day, R. S., & Bartlett, J. C. Separate speech and non-speech processing in dichotic listening? Journal of the Acoustical Society of America, 1972, 51, 79.

Derwing, B. Transformational grammar as a theory of language *acquisition: A study in the empirical, conceptual, and methodological foundations of contemporary linguistic theory.* Cambridge: Cambridge University Press, 1973.

Eimas, P. D., Ziqusland, E. R., Jusczyk, P., & Vigorito, J. Speech perception in infants. *Science*, 1971, *171*, 303-306.

Fant, G. *Speech sounds and features.* Cambridge Mass.: M.I.T. Press, 1973.

Fromkin, V. The concept of "naturalness" in a universal phonetic theory. *Glossa*, *4(1)*, 29-45.

Garner, W. R., & Felfoldy, G. L. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1970, *1(3)*, 225-241.

Garnica, O. K. The development of the perception of phonemic differences in initial consonants by English speaking children: A pilot study. *Papers and reports on child language development*, Stanford University, 1971, *3*, 1-30.

Graham, W., & House, A. S. Phonological opposition in young children: A perceptual study. *Journal of the Acoustical Society of America*, 1971, *49*, 559-566.

Halle, M. Phonology in a generative grammar. *Word*, 1962, *18*, 54-72.

Harman, H. H. *Modern factor analysis.* Chicago: University of Chicago Press, (2nd ed.), 1967.

Harms, R. T. *Introduction to phonological theory.* New Jersey: Prentice-Hall, 1968.

Harshman, R. Foundations of the PARAFAC proceedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 1970, *16*, 1-84.

Householder, F. W. On some recent claims in phonological theory. *Journal of Linguistics*, 1965, *1*, 13-34.

Hyman, R., & Well, A. Judgements of similarity and spatial models. *Perception and Psychophysics*, 1967, *2*, 233-248.

Hyman, R., & Well, A. Perceptual separability and spatial models. *Perception and Psychophysics*, 1968, *3*, 161-165.

Indow, T., & Uchizono, T. Multidimensional mapping of Munsell colors varying in hue and chroma. *Journal of Experimental Psychology*, 1960, *59*, 321-329.

Jakobson, R., & Halle, M. *Fundamentals of language.* The Hague: Mouton, 1956.

Jakobson, R., Fant, G. M., & Halle, M. *Preliminaries to speech analysis.* Cambridge Mass.: M.I.T. Press, 1951.

Jeter, I., & Singh, S. A comparison of phonemic and graphemic features of eight English consonants under auditory and visual modes. *Journal of Speech and Hearing Research*, 1972, *15*, 201-210.

Johnson, S. C. Hierachical clustering schemes. *Psychometrica* , 1967, *32*, 241-245.

Kimura, D. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, 1961, *15*, 166-171.

Klahr, D. A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling proceedure. *Psychometrika*, 1969, *34*, 319-333.

Kruskal, J. B. Multidemensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, 1964, *29(1)*, 1-27.

Kruskal, J. B. Nonmetric multidimensional scaling, a numerical method. *Psychometrika*, 1964, *29(2)*, 115-129.

Ladefoged, P. Phonological features and their phonetic correlates. *UCLA Working Papers in Phonetics*, 1971, *21*, 3-12.

Lane, H. L. A behavioural basis for the polarity principle in linguistics. *Language*, 1967, *43*, 494-511.

Lane, H. L. The motor theory of speech perception: A critical review. *Psychological Review*, 1965, *72*, 275-309.

Liberman, A. M. The grammars of speech and language. *Cognitive Psychology*, 1970, *1*, 301-323.

Liberman, A. M., Cooper, F. S., Shankweiler, D. S., & Studdert-Kennedy, M. Perception of the speech code. *Psycholoogical Review*, 1967, *74*, 431-461.

Liberman, A. M., Cooper, F. S., Harris, K. S., & MacNeilage, P. F. A motor theory of speech perception. In C. G. M. Fant (Ed.), *Proceedings of the speech communication seminar*, Stockholm, 1962.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 1957, 54, 358-368.

Liberman, A. M., Harris, K. S., Eimas, P., Lisker, L., & Bastion, J. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. Language and Speech, 1961, 4, 216 229.

Massaro, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. Psychological Review, 1972, 79, 124-125.

McGee, V.E. The multidimensional analysis of 'elastic' distances. British Journal of Mathematical and Statistical Psychology, 1966, 19, 181-196.

Messick, S. J., & Abelson, R. P. The addative constant problem in multidimensionalscaling. Psychometrika, 1956, 21, 1-15.

Miller, G. A. & Nicely, P. E. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 1955, 27, 338-352.

Mountcastle, V. B. Neural mechanisms in somesthesia. In V. B. Mountcastle (Ed.), Medical Physiology, Vol. 2., Saint Louis: Mosby, 1968.

Peters, R. W. Dimensions of perception for consonants. Journal of the Acoustical Society of America, 1963, 35, 1985-1989.

Pols, L. C. W. Intelligibility of speech resynthesised using a dimensional spectral representation. Paper presented at the Speech Communication Semminar, Stockholm, August, 1974.

Pols, L. C. W., vander Kamp, L. T. Th., & Plomp, R. Perceptual and physical space of vowel sounds. Journal of the Acoustical Society of America, 1969, 46, 458-467.

Postal, P. Aspects of phonological theory. New York: Harper and Row, 1968.

Pruzansky, S. Judgements of similarities among initial consonants using an auditory sorting apparatus. Journal of the Acoustical Society of America, 1971, 49, 84.

Roszypal, A. J. Computer supported gating of speech signals. Paper presented at the Speech Communication Semminar, Stockholm, August, 1974. Also forthcoming in Speech and Language, 1975.

Schane, S. A. *Generative phonology.* New Jersey: Prentice-Hall, 1973.

Shepard, R. N. Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, 1963, *5*, 33-48.

Shepard, R. N. Attention and the metric structure of the stimulus. *Journal of Mathematical Psychology*, 1964, *1*, 54-87.

Shepard, R. N. Psychological representation of speech sounds. In, E. E. David & P. B. Denes (Eds.), *Human communication: A unified view*, New York: McGraw-Hill, 1972.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 1962, *27*, 219-246.

Shepard, R. N., Kimball, A. R., & Nerlove, S. B., *Multidimensional scaling: Theory and applications in the behavioural sciences.* 2 Vols., New York: Semminar Press, 1972.

Shvachkin, N. Kh. The development of phonemic perception in early childhood. In, C. A. Ferguson, & D. I. Slobin (Eds.), *Studies in child language development.* New York: Holt, Rinehart, and Winston, 1973.

Singh, S. A step towards a theory of speech perception. Paper presented at the Speech Communication Seminar, Stockholm, Aug. 1-3, 1974.

Singh, S., & Black, J. W. Study of 26 intervocalic consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America*, 1966, *39*, 372-378.

Singh, S., Woods, D. R., & Becker, G. Perceptual structure of 22 prevocalic English consonants. *Jorunal of the Acoustical Society of America*, 1972, *52*, 1688-1712.

Singh, S., Woods, D. R., & Tishman, A. An alternative MD-SCAL analysis of the Graham and House data. *Journal of the Acoustical Society of America*, 1972, *51*, 666-668.

Smith, P. T. Feature testing models and their application to perception and memory for speech. *Quarterly Journal of Experimental Psychology*, 1973, *25*, 511-534.

Stanley, R. Redundancy rules in phonology. *Language*, 1967, *43*, 393-436.

Stanley, R. Boundaries in phonology. In S. R. Anderson and P. Kiparsky (Eds.), A festschrift for Morris Halle, New York: Holt, Rinehart and Winston, 1973.

Stenson, H. H., & Knoll, R. L. Goodness and badness of fit for random rankings in Kruskal's nonmetric scaling proceedure. Psychological Bulletin, 1969, 71, 122-126.

Studdert-Kennedy, M., & Shankweiler, D. Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 1970, 48, 579 -591.

Terbeek, D., & Harshman, R. Cross-language differences in the perception of natural vowel sounds. UCLA Working Papers in Phonetics, 1971, 19, 26-38.

Thurstone, L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Torgerson, W. S. Multidimensional scaling of similarity. Psychometrika, 1965, 30, 379-393.

Torgerson, W. S. Multidimensional scaling: I - theory and method. Psychometrika, 1952, 17, 401-419.

Torgerson, W. S. Theory and Methods of Scaling. New York: Wiley, 1958.

Trubetzkoy, N. S. Principles of Phonology. Berkeley: University of California Press, 1969.

Veldman, D. J. Fortran programming for the behavioural sciences. New York: Holt, Rinehart, and Winston, 1967.

Vennemann, T., & Ladefoged, P. Phonetic features and phonological features. UCLA Working Papers in Phonetics, 1971, 21, 13-24.

Wang, M., & Bilger, R. G. Consonant confusions in noise: A study of perceptual features. Jorurnal of the Acoustical Society of America, 1973, 54, 1248-1266.

Whitfield, I. C., & Evans, E. F. Responses of auditory cortical neurons to stimuli of changing frequency. Journal of Neurophysiology, 1965, 28, 655-672.

Wickelgren, W. A. Distinctive features and errors in short-term memory for English consonants. Journal of the Acoustic Society of America, 1966, 39(2), 388-398.

Wilson, K. V. Multidimensional analysis of confusions of English consonants. American Journal of Psychology, 1963, 76, 89-95.

Wish, M. An INDSCAL analysis of the Miller and Nicely consonant confusion data. Paper presented at the Acoustical Society of America meeting, Houston, 1970.

Wood, C. C. Levels of processing in speech perception: Neurophysiological and information-processing analyses. Thesis Supplement in _Speech Research_, Haskins Laboratories, 1973, _35/36_.

Wood, C. C., Goff, W. R., & Day, R. S. Auditory evoked potentials during speech perception. _Science_, 1971, _173_, 1248-1251.

Worden, F. G., & Galambos, R. Auditory processing of biologically significant sounds. _Neurosciences Research Program Bulletin_, 1972, _10(1)_, 1-119.

Young, F. W. Nonmetric multidimensional scaling: Recovery of metric information. _Psychometrika_, 1970, _35(4)_, 455-473.

Young, F. W., & Torgerson, W. S. TORSCA, a FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. _Behavioural Science_, 1967, _12_, 498.

APPENDIX A
DISTINCTIVE FEATURE DEFINITIONS

Anterior: "sounds are produced with an obstruction that is located in front of the palatoalveolar region of the mouth [Chomsky and Halle, 1967, 304]. This feature, stricktly speaking, applies only to consonantal sounds.

Compact (vs. Diffuse): 'Compact phonemes are characterized by the relative predominance of one centrally located formant region [Jakobson, Fant, and Halle (hereafter, JFH), 1951, 27]." This feature co-classifies consonantal sounds made with a constriction in the posterior portion of the oral cavity (velars, palatals) and open vowels.

Continuant: If air-flow through the mouth is not blocked during production of a sound then such a sound is labelled continuant. The liquids [l] and [r] are difficult to classify on this dichotomous scale.

Consonantal: JFH define this feature in acoustic terms "by the presence of zeros that affect the entire spectrum [p.19]." C&H define it articulatorily as those sounds "produced in the midsaggital region of the [oral] cavity [p.302]." In either case, the close parallel with the feature Vocalic is obvious and it is doubtful whether these two features represent distinct dimensions for the native speaker or listener.

Coronal: All sounds produced with the blade of the tongue raised above the neutral position are labelled "coronal". This feature distinguishes dentals, alveolars and alveopalatal sounds from labials, velars, palatals, and pharyngeals. It is, stricktly speaking, a consonantal feature.

Frication: This feature characterizes all sounds with a non-plosive turbulent noise component.

Grave (vs. Acute): "This feature means the predominance of one side of the significant part of the spectrum over the other. When the lower side of the spectrum predominates, we term the phoneme acute [JFH, p.29]." This feature co-classifies labial and velar consonants and vowels with a high second formant. JFH admit that a complex normalization of the signal would be necessary to achieve automatic separation of speech sounds on this hypothetical dimension.

High: sounds are produced with the tongue body elevated above the "neutral" position. Velar and palatal consonants are regarded as "high", as are the traditional high vowels.

Low: sounds are produced with the body of the tongue below the neutral position. The traditional low vowels and pharyngeal and glottal consonants are regarded as "low".

Nasal: sounds are characterized by a lowered velum, with or without closure of the oral cavity. The acoustic coupling of the nasal cavity introduces additional poles and zeros into the supraglottal transfer function.

Place (SB): a four valued categorical place of articulation feature used by Singh and Black (1968), classifying sounds into (1)labials, (2)dentals and alveolars, (3)palatals, (4)velars.

Sibilant: sibilants are sounds posessing the greatest amount of turbulent noise. This feature separates the alveolar and alveopalatal fricatives from the softer labiodental fricatives and all other noise-weak sounds.

Strident: "sounds are marked by greater noisiness than their non-strident counterparts [C&H, p.329]." For C&H the relatively weak frication in the English labiodental fricatives is sufficiently strong to be "strident". JFH on the other hand regard the labiodental fricatives as non-strident.

Tense (vs. Lax): "Tense sounds are produced with a deliberate, accurate, maximally distinct gesture that involves considerable muscular effort [C&H,p.324]." This feature differentiates voiceless from voiced consonants and vowels according to their degree of constriction (amount of movement of the tongue body from the neutral position).

Voice: The voicing feature has been variously defined in articulatory and acoustic terms. Most simply it is characterized by the presence of glottal activity up to the point of maximal constriction in the articulation of the sound. The presence of a spectral voice bar and voice onset time (VOT) are often treated as defining characteristics of this feature. With respect to the consonants, this feature is co-extensive with the tense-lax distinction.

Vocalic: Vocalic sounds posess a "single periodic voice source whose onset is not abrupt [JFH, p.18]." This feature is used to distinguish vowel and vowel-like sounds from consonantal type sounds. C&H characterize this feature in terms of degree of constriction of the oral cavity. Formant structure is an important accompanying, but not defining, characteristic of this feature.

APPENDIX B
KRUSKAL SCALING EXPERIMENT I

COORDINATES FOR 3-DIMENSIONAL SOLUTION
STRESS = .05

	DI	DII	DIII
pa	0.197	0.548	0.841
ba	0.051	0.793	0.405
ta	-0.789	-0.080	0.817
da	-0.092	1.127	-0.034
ca	-0.776	-0.809	0.026
sa	-0.458	-0.776	-0.790
sa	-0.517	-1.022	-0.366
ha	0.315	-0.286	0.383
za	-0.702	-0.392	-0.979
ma	0.836	0.485	-0.098
na	0.740	0.502	-0.204
la	1.104	0.134	-0.029

COORDINATES FOR 2-DIMENSIONAL SOLUTION
STRESS = .10

	DI	DII
pa	-0.932	0.131
ba	-0.754	0.291
ta	-0.734	-0.800
da	-0.715	0.549
ca	0.331	-1.041
sa	1.007	-0.703
sa	0.811	-0.883
ha	-0.047	0.064
za	1.170	-0.456
ma	-0.159	0.987
na	-0.085	0.962
la	0.107	0.897

APPENDIX C
RAW PROXIMITY MATRICES EXPERIMENT II

/Ca/ set

	sa	pa	ca	ma	ta	la	da	ha	sa	za	na	ba
sa												
pa	106											
ca	87	112										
ma	110	60	129									
ta	110	58	64	97								
la	93	87	110	32	88							
da	100	77	86	67	50	67						
ha	89	25	95	53	88	50	96					
sa	28	95	24	99	117	103	98	73				
za	4	123	89	113	103	106	92	102	64			
na	96	87	92	0	87	46	56	66	112	99		
ba	96	24	126	54	79	79	25	58	133	115	67	

/Ci/ set

	si	pi	ci	mi	ti	li	di	hi	si	zi	ni	bi
si												
pi	102											
ci	93	114										
mi	129	106	148									
ti	108	71	134	121								
li	123	98	125	62	103							
di	108	51	134	113	44	97						
hi	111	78	111	80	104	64	92					
si	83	129	30	126	134	128	141	93				
zi	35	138	90	123	125	117	110	110	90			
ni	122	110	138	0	107	68	97	89	128	123		
bi	121	53	133	77	93	104	64	109	138	121	69	

APPENDIX D
KRUSKAL SCALING EXPERIMENT II

/Ca/ set

COORDINATES FOR
3-DIMENSIONAL SOLUTION
STRESS = .08

	DI	DII	DIII
sa	0.854	0.561	-0.476
pa	-0.480	-0.731	-0.281
ča	1.034	-0.443	0.182
ma	-0.929	0.210	-0.245
ta	0.100	-0.646	0.795
la	-0.609	0.572	-0.164
da	-0.238	-0.144	0.758
ha	-0.338	-0.142	-0.663
ša	0.974	-0.027	-0.631
za	0.998	0.833	0.113
na	-0.576	0.448	0.390
ba	-0.791	-0.446	0.223

COORDINATES FOR
2-DIMENSIONAL SOLUTION
STRESS = .14

	DI	DII
sa	1.017	0.648
pa	-0.731	-0.431
ča	1.079	-0.475
ma	-0.877	0.250
ta	0.147	-1.100
la	-0.842	0.467
da	-0.086	-0.749
ha	-0.468	0.260
ša	1.118	0.216
za	0.958	0.939
na	-0.669	0.548
ba	-0.646	-0.573

/Ci/ set

STRESS = .06

	DI	DII	DIII
si	0.877	-0.437	0.243
pi	-0.659	-0.530	-0.174
či	1.141	-0.075	-0.613
mi	-0.475	0.895	0.230
ti	-0.476	-0.898	0.200
li	-0.563	0.608	-0.365
di	-0.545	-0.545	0.177
hi	-0.159	0.250	-0.557
ši	1.105	0.245	-0.532
zi	1.039	-0.056	0.408
ni	-0.573	0.630	0.367
bi	-0.708	-0.087	0.617

STRESS = .09

	DI	DII
si	1.017	0.648
pi	-0.731	-0.431
či	1.079	-0.475
mi	-0.877	0.250
ti	0.147	-1.100
li	-0.842	0.250
di	-0.086	-0.749
hi	-0.468	0.260
ši	1.118	0.216
zi	0.958	0.939
ni	-0.669	0.548
bi	-0.646	-0.573

```
23:39.23 37.122 RC=0
```

TORGERSON SCALING EXPERIMENT II

ABSOLUTE DISTANCE MATRIX /Ca/ SET

	ba	sa	da	ha	la	ma	na	pa	sa	ca	ta	za
ba												
sa	0.707											
da	-0.694	0.219										
ha	-0.373	-0.077	0.056									
la	-0.052	0.375	-0.219	-0.609								
ma	-0.428	0.101	-0.386	-0.493	-0.705							
na	-0.236	0.359	-0.456	-0.300	-0.565	-1.236						
pa	-0.690	0.299	-0.121	-0.761	0.073	-0.319	-0.130					
sa	0.221	-0.634	0.277	0.020	0.203	0.303	0.095	0.296				
ca	0.604	-0.516	0.275	0.256	0.402	0.607	0.298	0.466	0.229			
ta	-0.039	0.515	-0.411	-0.016	0.036	0.194	0.067	-0.263	0.453	-0.049		
za	0.497	-0.058	0.269	0.317	0.405	0.453	0.294	0.630	-0.805	0.372	0.405	

APPENDIX E

APPENDIX E (CONTINUED)
PROJECTION OF STIMULI IN 4 DIMENSIONS

	I	II	III	IV
ba	-0.7351	-0.0934	-0.3771	-0.3356
sa	0.9830	0.0090	0.5079	-0.1008
da	-0.3625	0.1825	-0.4051	0.2966
ha	-0.2321	-0.1235	0.4493	-0.3425
la	-0.3948	-0.2396	0.2577	0.4869
ma	-0.5466	-0.4090	0.3382	0.2091
na	-0.3841	-0.2456	0.0654	0.4396
pa	-0.5577	0.1844	0.1392	-0.6671
sa	0.8130	-0.5054	-0.1951	-0.2523
ca	0.7688	0.8837	0.2515	0.2045
ta	-0.2524	0.8151	-0.4033	0.0206
za	0.9005	-0.4580	-0.6288	0.0409

APPENDIX E CONTINUED

TORGERSON SCALING EXPERIMENT II

ABSOLUTE DISTANCE MATRIX /Ci/ SET

	ba	sa	da	ha	la	ma	na	pa	sa	ca	ta	za
ba												
sa	0.609											
da	-0.514	0.648										
ha	0.004	0.015	-0.230									
la	-0.043	0.428	-0.161	-0.623								
ma	-0.325	0.304	-0.024	-0.507	-0.663							
na	-0.379	0.353	-0.075	-0.381	-0.428	-1.540						
pa	-0.649	0.460	-0.596	-0.436	-0.172	-0.059	-0.117					
sa	0.325	-0.117	0.144	0.053	0.281	0.344	0.248	0.010				
ca	0.568	-0.695	0.391	0.146	0.388	0.655	0.435	0.291	0.010			
ta	-0.088	0.477	-0.602	-0.003	0.051	0.196	-0.036	-0.390	0.279	0.352		
za	0.287	0.014	0.219	0.009	0.307	0.280	0.282	0.351	-0.698	-0.015	0.349	

APPENDIX E (CONTINUED)

PROJECTION OF STIMULI IN 4 DIMENSIONS

	I	II	III	IV
ba	-0.5762	-0.1434	-0.1701	0.3966
sa	0.9162	0.3157	0.2562	0.1948
da	-0.4630	-0.4450	-0.0357	-0.1521
ha	-0.0719	0.2929	0.1084	-0.2553
la	-0.3704	0.3879	0.1028	-0.3236
ma	-0.4307	0.5774	-0.1413	0.0649
na	-0.3691	0.3159	0.0262	0.2055
pa	-0.3696	-0.3191	0.1241	-0.0052
sa	0.5775	-0.2239	-0.4325	-0.0816
ca	0.8716	-0.1699	0.4279	0.0268
ta	-0.3168	-0.5069	0.2956	-0.0106
za	0.6023	-0.0815	-0.5617	-0.0603

APPENDIX E CONTINUED

APPENDIX F

PROXIMITY MATRIX ROW CORRELATIONS /Ca/ SET

EXPERIMENT II

	sa	pa	ca	ma	ta	la	da	ha	sa	za	na
pa	-0.558										
ca	0.170	-0.499									
ma	-0.539	0.492	-0.550								
ta	-0.694	0.318	-0.059	0.062							
la	-0.385	0.220	-0.519	0.749	0.008						
da	-0.559	0.344	-0.298	0.472	0.493	0.388					
ha	-0.392	0.791	-0.472	0.666	-0.054	0.497	0.064				
sa	0.671	-0.403	0.679	-0.566	-0.595	-0.546	-0.602	-0.236			
za	0.914	-0.646	-0.002	-0.536	-0.493	-0.346	-0.446	-0.519	0.410		
na	-0.463	0.224	-0.448	0.932	0.086	0.789	0.433	0.513	-0.555	-0.384	
ba	-0.566	-0.566	0.725	-0.704	0.676	0.434	0.425	0.728	0.514	-0.754	0.555

APPENDIX F

APPENDIX F CONTINUED

UNROTATED PRINCIPAL AXES FACTORS
/Ca/ SET EXPERIMENT II

	COMMUNALITIES	I	II	III	IV
sa	0.944	-0.792	0.485	-0.157	-0.239
pa	0.981	0.716	-0.086	0.515	-0.443
ca	0.939	-0.631	-0.420	0.273	0.538
ma	0.934	0.856	0.380	0.024	0.239
ta	0.855	0.459	-0.781	-0.174	-0.061
la	0.831	0.690	0.449	-0.184	0.346
da	0.704	0.672	-0.317	-0.387	0.036
ha	0.959	0.657	0.389	0.600	-0.123
sa	0.877	-0.809	0.154	0.426	0.131
za	0.971	-0.706	0.402	-0.462	-0.312
na	0.926	0.756	0.399	-0.173	0.407
ba	0.912	0.873	-0.037	-0.137	-0.359
		6.326	1.973	1.384	1.149
PERCENT OF COMMON VARIANCE		58.398	18.218	12.747	10.611
PERCENT OF TOTAL VARIANCE	90.267	52.714	16.445	11.530	9.578

VARIMAX ROTATED FACTORS

	I	II	III	IV
sa	-0.818	-0.318	-0.414	0.056
pa	0.294	0.009	0.872	0.366
ca	0.018	-0.311	-0.217	-0.891
ma	0.178	0.831	0.383	0.253
ta	0.905	-0.149	0.032	0.108
la	0.008	0.874	0.108	0.218
da	0.699	0.315	-0.045	0.338
ha	-0.098	0.400	0.869	0.187
sa	-0.626	-0.386	-0.035	-0.579
za	-0.653	-0.300	-0.620	0.263
na	0.170	0.921	0.130	0.180
ba	0.479	0.295	0.374	0.675
PERCENT OF COMMON VARIANCE	29.455	28.690	22.499	19.356

PROXIMITY MATRIX ROW CORRELATIONS /Ci/ SET

EXPERIMENT II

	sa	pa	ca	ma	ta	la	da	ha	sa	za	na
pa	-0.375										
ca	0.232	-0.417									
ma	-0.615	0.208	-0.673								
ta	-0.214	0.621	-0.547	0.014							
la	-0.549	0.285	-0.578	0.665	0.171						
da	-0.258	0.760	-0.574	0.071	0.864	0.222					
ha	-0.353	0.308	-0.309	0.398	0.118	0.657	0.148				
sa	0.348	-0.561	0.854	-0.511	-0.699	-0.588	-0.740	-0.257			
za	0.854	-0.544	0.162	-0.439	-0.361	-0.41[illegible]	-0.343	-0.260	0.286		
na	-0.521	0.115	-0.706	0.957	0.013	0.680	0.092	0.381	-0.548	-0.353	
ba	-0.463	0.723	-0.710	0.552	0.476	0.328	0.651	0.123	-0.747	-0.396	0.521

APPENDIX F CONTINUED

UNROTATED PRINCIPAL AXES FACTORS

/Ci/ SET EXPERIMENT II

	COMMUNALITIES	I	II	III	IV
sa	0.915	-0.667	0.313	0.583	0.180
pa	0.792	0.694	0.473	-0.284	0.073
ca	0.953	-0.825	-0.030	-0.521	0.011
ma	0.960	0.733	-0.596	0.139	-0.221
ta	0.843	0.604	0.680	-0.019	0.128
la	0.850	0.719	-0.443	0.028	0.368
da	0.926	0.682	0.674	-0.004	0.078
ha	0.907	0.477	-0.341	-0.124	0.740
sa	0.908	-0.866	-0.232	-0.322	-0.012
za	0.934	-0.617	0.087	0.720	0.165
na	0.961	0.711	-0.592	0.262	-0.189
ba	0.842	0.815	0.217	0.106	-0.347
		6.020	2.365	1.430	0.975
PERCENT OF COMMON VARIANCE		55.788	21.920	13.251	9.041
PERCENT OF TOTAL VARIANCE	89.918	50.164	19.710	11.915	8.129

VARIMAX ROTATED FACTORS

	I	II	III	IV
sa	-0.116	-0.328	-0.871	-0.191
pa	0.786	0.022	0.394	0.137
ca	-0.569	-0.762	0.127	-0.179
ma	-0.018	0.892	0.350	0.205
ta	0.912	-0.019	0.075	0.068
la	0.137	0.552	0.239	0.685
da	0.954	0.057	0.108	0.048
ha	0.090	0.149	0.147	0.925
sa	-0.740	-0.578	-0.010	-0.162
za	-0.253	-0.110	-0.917	-0.126
na	-0.018	0.929	0.226	0.215
ba	0.649	0.565	0.280	-0.150
	3.756	3.343	2.119	1.572

APPENDIX G

UNROTATED PRINCIPAL AXES FACTORS
EXPERIMENT IV

	COMMUNALITIES	I	II
pa	0.850	0.887	-0.251
ma	0.951	-0.060	0.974
ca	0.865	-0.139	-0.919
ba	0.796	0.874	0.182
sa	0.933	-0.965	-0.046
ta	0.808	0.715	-0.545
na	0.934	0.031	0.966
za	0.891	-0.752	-0.570
la	0.771	-0.046	0.877
ha	0.416	-0.242	0.598
sa	0.658	-0.752	-0.304
da	0.938	0.954	-0.170
		5.119	4.694

PERCENT OF COMMON VARIANCE

		I	II
		52.166	47.834

PERCENT OF TOTAL VARIANCE

		I	II
	81.776	42.659	39.117

VARIMAX ROTATED FACTORS

	I	II
pa	0.872	-0.301
ma	-0.005	0.975
ca	-0.191	-0.910
ba	0.883	0.132
sa	-0.966	0.009
ta	0.683	-0.584
na	0.086	0.963
za	-0.784	-0.526
la	0.004	0.878
ha	-0.208	0.611
sa	-0.768	-0.260
da	0.942	-0.224

PERCENT OF COMMON VARIANCE

	I	II
	52.152	47.848

APPENDIX H
PHONOLOGICAL FEATURES: REGRESSION ANALYSES

PREDICTORS: Jakobson, Fant, and Halle features
CRITERION: Kruskal interpoint distances

FEATURES	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Strident-M.	0.780	0.608	0.608	0.761
Consonantal-V.	0.789	0.622	0.014	-0.097
Continuant-I.	0.799	0.639	0.017	0.131
Grave-Acute	0.801	0.642	0.003	0.055
Compact-Diffuse	0.803	0.644	0.002	-0.062
Tense-Lax	0.805	0.647	0.003	0.053
Nasal-Non nasal	0.805	0.647	0.001	0.024

PREDICTORS: Jakobson, Fant, and Halle features
CRITERION: Raw proximity scores

FEATURES	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Strident-M.	0.715	0.511	0.511	0.676
Tense-Lax	0.743	0.552	0.041	0.197
Continuant-I.	0.763	0.582	0.030	0.171
Grave-Acute	0.773	0.597	0.015	0.122
Nasal-Non nasal	0.775	0.602	0.004	0.076
Compact-Diffuse	0.777	0.604	0.002	0.088
Consonantal-V.	0.780	0.608	0.004	-0.068
Vocalic-N.	0.781	0.611	0.002	0.052

PREDICTORS: Chomsky and Halle features
CRITERION: Kruskal interpoint distances

FEATURES	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Strident .	0.780	0.608	0.608	0.745
Low	0.789	0.621	0.014	-0.134
Continuant	0.799	0.639	0.017	0.145
Coronal	0.805	0.649	0.010	0.098
Voiced-Vcls.	0.806	0.651	0.003	0.058
Anterior	0.808	0.653	0.002	-0.047
Nasal	0.809	0.654	0.001	0.026

PREDICTORS: Chomsky and Halle features
CRITERION: Raw proximity scores

FEATURES	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Strident	0.715	0.511	0.511	0.637
Voiced-Vcls.	0.743	0.552	0.041	0.210
Coronal	0.767	0.589	0.036	0.252
Continuant	0.789	0.624	0.035	0.206
Low	0.798	0.637	0.013	-0.160
Anterior	0.804	0.647	0.010	0.127
Nasal	0.807	0.652	0.004	0.077

APPENDIX H (CONTINUED)

PREDICTORS: Singh and Black features
CRITERION: Kruskal interpoint distances

FEATURES	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Duration	0.598	0.358	0.358	0.480
Frication	0.674	0.454	0.096	0.336
Nasal	0.696	0.485	0.031	0.264
Place	0.721	0.520	0.034	-0.205
Voice-Vcls.	0.721	0.520	0.000	0.016

PREDICTORS: Singh and Black features
CRITERION: Raw proximity scores

FEATURES	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Frication	0.552	0.305	0.305	0.356
Duration	0.635	0.403	0.098	0.400
Nasal	0.660	0.435	0.032	0.249
Voiced-Vcls.	0.684	0.468	0.032	0.180

PREDICTORS: Ingram features (set #1)
CRITERION: Kruskal interpoint distances

FEATURES	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Duration	0.640	0.409	0.409	0.548
Resonance-Hiss	0.841	0.708	0.298	0.655
Nasal	0.859	0.739	0.031	-0.202
Voiced-Vcls.	0.861	0.741	0.001	-0.051
Place	0.862	0.743	0.001	0.043

PREDICTORS: Ingram features (set #2)
CRITERION: Kruskal interpoint distances

FEATURES	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Sibilance	0.779	0.608	0.608	0.759
Resonance	0.822	0.676	0.067	0.266
Stop	0.859	0.738	0.062	0.249
Voiced-Vcls.	0.860	0.740	0.002	0.048
Nasal	0.860	0.741	0.001	-0.035

APPENDIX I
RATING SCALES: REGRESSION ANALYSIS

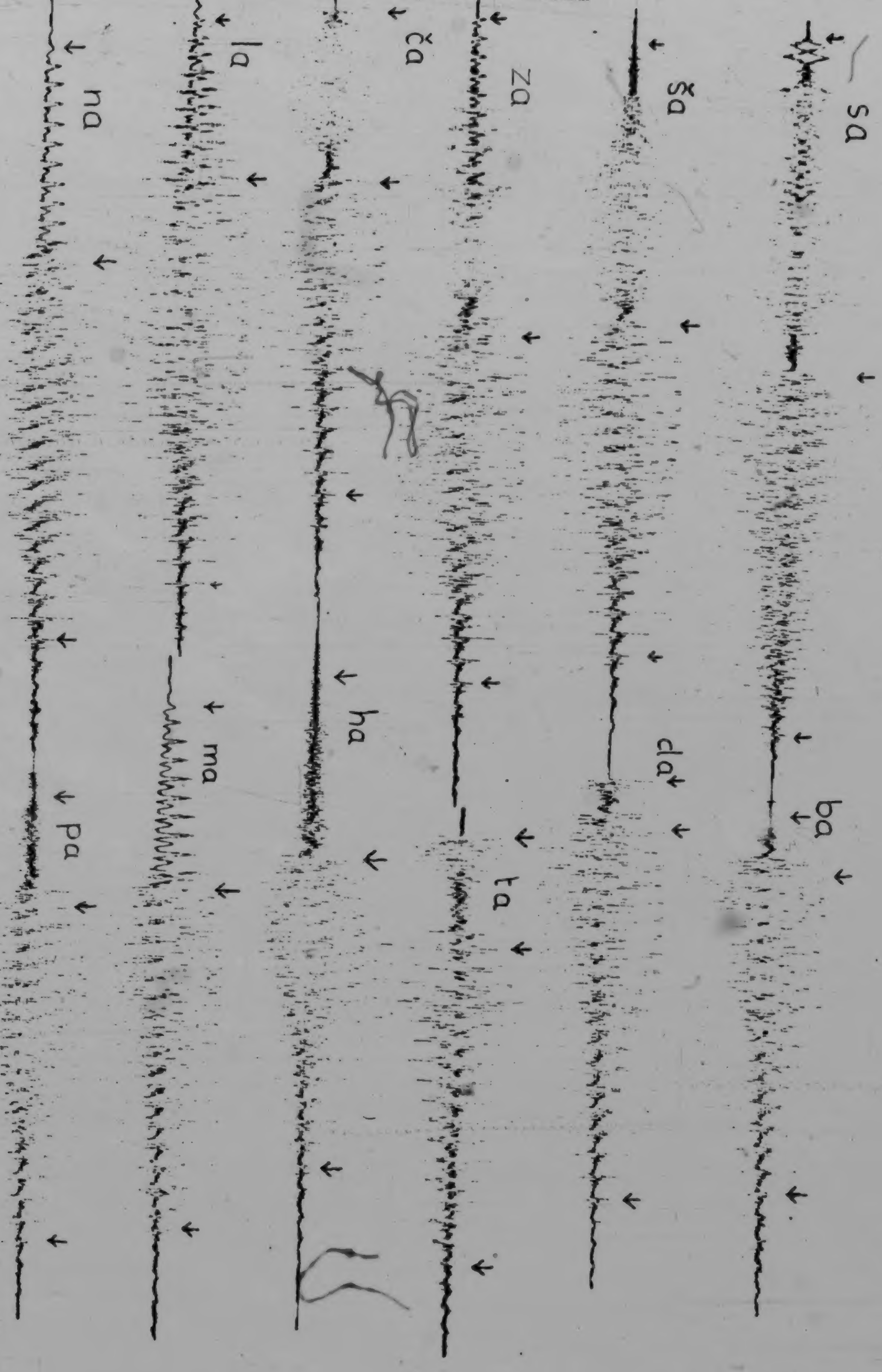
PREDICTORS: Verbal rating scales of sound quality
CRITERION: Kruskal interpoint distances

SCALE	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Hissy	0.762	0.581	0.581	0.761
Abrupt	0.846	0.715	0.305	0.216
Vowel-like	0.882	0.778	0.063	0.333
Distinct	0.903	0.815	0.036	0.226
Harsh	0.906	0.822	0.006	0.255
Melodious	0.914	0.836	0.014	-0.217
Loud	0.917	0.840	0.003	-0.078
Short-Long	0.917	0.841	0.001	0.136
Clear	0.919	0.845	0.003	-0.130

PREDICTORS: Verbal rating scales of sound quality
CRITERION: Raw proximity scores

SCALE	MULTIPLE R	R SQUARE	RSQ CHANGE	BETA
Hissy	0.697	0.486	0.486	0.587
Abrupt	0.757	0.573	0.086	0.126
Harsh	0.780	0.608	0.035	0.091
Distinct	0.792	0.628	0.020	0.152
High-Pitch	0.803	0.646	0.018	0.156
Vowel-like	0.809	0.655	0.010	0.123
Even	0.812	0.659	0.003	0.081
Short-Long	0.813	0.660	0.002	0.077

APPENDIX J
OSCILLOGRAMS OF TEST STIMULI



APPENDIX K
ACOUSTIC PROPERTIES: REGRESSION ANALYSES

```
PREDICTORS:  BPGP - High (band-pass) filter
                    peak amplitude, log. scale.
             LPGP - Low-pass filter peak amplitude
                    log. scale.
CRITERION:   Loadings on Kruskal "Resonance - Hiss"
                    factor.

  VARIABLE    MULTIPLE R   R SQUARE   RSQ CHANGE    BETA
BPGP            0.857        0.734      0.734      -0.755
LPGP            0.913        0.834      0.100       0.333
```

```
PREDICTOR:   CDUR - Duration of consonant
CRITERION:   Loadings on Kruskal "Duration" factor

  VARIABLE     SIMPLE R   R SQUARE     B      BETA
CDUR            0.954       0.911    2.553   0.954
                  CONSTANT =  -12.995
```

APPENDIX L
BANDFILTER FUNCTIONS OF EXPERIMENTAL STIMULI